

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ :		A2	(11) International Publication Number:	WO 00/20601
C12N 15/52, 15/54, 15/62, 9/10, C12P 17/18, 19/32, C07D 498/18 // (C07D 498/18, 311:00, 273:00, 211:00)			(43) International Publication Date:	13 April 2000 (13.04.00)
(21) International Application Number:		PCT/US99/22886	(74) Agents: FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).	
(22) International Filing Date:		1 October 1999 (01.10.99)	(81) Designated States: AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, IL, IS, JP, KG, KP, KR, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data:			Published <i>Without international search report and to be republished upon receipt of that report.</i>	
60/102,748 60/123,810 60/139,650		2 October 1998 (02.10.98) 11 March 1999 (11.03.99) 17 June 1999 (17.06.99)	US US US	
(71) Applicant (for all designated States except US): KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).				
(72) Inventors; and				
(75) Inventors/Applicants (for US only): REEVES, Christopher [US/US]; 4 East Altarinda Drive, Orinda, CA 94563 (US). CHU, Daniel [US/US]; 3767 Benton Street, Santa Clara, CA 95051 (US). KHOSLA, Chaitan [IN/US]; 740 Para Avenue, Palo Alto, CA 94306 (US). SANTI, Daniel [US/US]; 211 Belgrave Avenue, San Francisco, CA 94117 (US). WU, Kai [CN/US]; 900 Constitution Drive, Foster City, CA 94404 (US).				
(54) Title: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR				
(57) Abstract				
<p>Host cells comprising recombinant vectors encoding the FK-520 polyketide synthase and FK-520 modification enzymes can be used to produce the FK-520 polyketide. Recombinant DNA constructs comprising one or more FK-520 polyketide synthase domains, modules, open reading frames, and variants thereof can be used to produce recombinant polyketide synthases and a variety of different polyketides with application as pharmaceutical and veterinary products.</p>				

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
CONSTRUCTS THEREFOR**

5

Field of the Invention

The present invention relates to polyketides and the polyketide synthase (PKS) enzymes that produce them. The invention also relates generally to genes encoding PKS enzymes and to recombinant host cells containing such genes and in which expression of such genes leads to the production of polyketides. The present invention also relates to 10 compounds useful as medicaments having immunosuppressive and/or neurotrophic activity. Thus, the invention relates to the fields of chemistry, molecular biology, and agricultural, medical, and veterinary technology.

15

Background of the Invention

Polyketides are a class of compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. Polyketides are biologically active molecules with a wide variety of structures, and the class encompasses numerous compounds with diverse activities. Tetracycline, 20 erythromycin, epothilone, FK-506, FK-520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of polyketides. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds.

25

This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce 30 molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters. See, e.g., PCT publication Nos. WO 93/13663; 95/08548; 96/40968; 97/02358; 98/27203; and 98/49315; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and Fu *et al.*, 1994, *Biochemistry* 33: 9321-9326; McDaniel *et al.*, 1993, 35 *Science* 262: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by PKS enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters, such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKS enzymes that incorporate such building blocks include an activity that functions as an amino acid ligase (an AMP ligase) or as a non-ribosomal peptide synthetase (NRPS). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesized. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

In the Type I or modular PKS enzyme group, a set of separate catalytic active sites (each active site is termed a "domain", and a set thereof is termed a "module") exists for each cycle of carbon chain elongation and modification in the polyketide synthesis pathway. The typical modular PKS is composed of several large polypeptides, which can be segregated from amino to carboxy termini into a loading module, multiple extender modules, and a releasing (or thioesterase) domain. The PKS enzyme known as 6-deoxyerythronolide B synthase (DEBS) is a Type I PKS. In DEBS, there is a loading module, six extender modules, and a thioesterase (TE) domain. The loading module, six extender modules, and TE of DEBS are present on three separate proteins (designated DEBS-1, DEBS-2, and DEBS-3, with two extender modules per protein). Each of the DEBS polypeptides is encoded by a separate open reading frame (ORF) or gene; these genes are known as *eryAII*, *eryAIII*, and *eryAI*. See Caffrey *et al.*, 1992, *FEBS Letters* 304: 205, and U.S. Patent No. 5,824,513, each of which is incorporated herein by reference.

Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The loading module of DEBS consists of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. Another type of loading module utilizes an inactivated ketosynthase (KS) domain and AT and ACP domains. This inactivated KS is in some instances called KS^Q , where the superscript letter is the abbreviation for the amino acid, glutanine, that is present instead of the active site cysteine required for ketosynthase activity. In other PKS enzymes, including the FK-506 PKS, the loading module

incorporates an unusual starter unit and is composed of a CoA ligase like activity domain. In any event, the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl or other acyl-CoA) and transfers it as a thiol ester to the ACP of the loading module.

5 The AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and 2-hydroxymalonyl) and transfers it to the ACP of that extender module to form a thioester. Each extender module is responsible for accepting a compound from a prior module, binding a building block, attaching the building block to the compound from the prior
10 module, optionally performing one or more additional functions, and transferring the resulting compound to the next module.

Each extender module of a modular PKS contains a KS, AT, ACP, and zero, one, two, or three domains that modify the beta-carbon of the growing polyketide chain. A typical (non-loading) minimal Type I PKS extender module is exemplified by extender
15 module three of DEBS, which contains a KS domain, an AT domain, and an ACP domain. These three domains are sufficient to activate a 2-carbon extender unit and attach it to the growing polyketide molecule. The next extender module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next extender module until synthesis is complete.

20 Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module is transferred to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module one possesses an acyl-KS and a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading building block (elongation or extension).

25 The polyketide chain, growing by two carbons each extender module, is sequentially passed as covalently bound thiol esters from extender module to extender module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly, however, additional enzymatic activities modify the beta keto group of each two carbon unit just after it has been added to the growing polyketide chain but before it is transferred to the next module.

Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, and as noted above, other domains that modify the beta-carbonyl moiety can be present. Thus, modules may contain a ketoreductase (KR) domain that reduces the keto group to an alcohol. Modules may also 5 contain a KR domain plus a dehydratase (DH) domain that dehydrates the alcohol to a double bond. Modules may also contain a KR domain, a DH domain, and an enoylreductase (ER) domain that converts the double bond product to a saturated single bond using the beta carbon as a methylene function. An extender module can also contain other enzymatic activities, such as, for example, a methylase or dimethylase 10 activity.

After traversing the final extender module, the polyketide encounters a releasing domain that cleaves the polyketide from the PKS and typically cyclizes the polyketide. For example, final synthesis of 6-dEB is regulated by a TE domain located at the end of extender module six. In the synthesis of 6-dEB, the TE domain catalyzes cyclization of 15 the macrolide ring by formation of an ester linkage. In FK-506, FK-520, rapamycin, and similar polyketides, the TE activity is replaced by a RapP (for rapamycin) or RapP like activity that makes a linkage incorporating a pipecolate acid residue. The enzymatic activity that catalyzes this incorporation for the rapamycin enzyme is known as RapP, encoded by the *rapP* gene. The polyketide can be modified further by tailoring enzymes; 20 these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated at C-6 and C-12 and glycosylated at C-3 and C-5 in the synthesis of erythromycin A.

In Type I PKS polypeptides, the order of catalytic domains is conserved. When 25 all beta-keto processing domains are present in a module, the order of domains in that module from N-to-C-terminus is always KS, AT, DH, ER, KR, and ACP. Some or all of the beta-keto processing domains may be missing in particular modules, but the order of the domains present in a module remains the same. The order of domains within modules is believed to be important for proper folding of the PKS polypeptides into an active 30 complex. Importantly, there is considerable flexibility in PKS enzymes, which allows for the genetic engineering of novel catalytic complexes. The engineering of these enzymes is achieved by modifying, adding, or deleting domains, or replacing them with those taken from other Type I PKS enzymes. It is also achieved by deleting, replacing, or adding entire modules with those taken from other sources. A genetically engineered

PKS complex should of course have the ability to catalyze the synthesis of the product predicted from the genetic alterations made.

Alignments of the many available amino acid sequences for Type I PKS enzymes has approximately defined the boundaries of the various catalytic domains. Sequence 5 alignments also have revealed linker regions between the catalytic domains and at the N- and C-termini of individual polypeptides. The sequences of these linker regions are less well conserved than are those for the catalytic domains, which is in part how linker regions are identified. Linker regions can be important for proper association between domains and between the individual polypeptides that comprise the PKS complex. One 10 can thus view the linkers and domains together as creating a scaffold on which the domains and modules are positioned in the correct orientation to be active. This organization and positioning, if retained, permits PKS domains of different or identical substrate specificities to be substituted (usually at the DNA level) between PKS enzymes by various available methodologies. In selecting the boundaries of, for example, an AT 15 replacement, one can thus make the replacement so as to retain the linkers of the recipient PKS or to replace them with the linkers of the donor PKS AT domain, or, preferably, make both constructs to ensure that the correct linker regions between the KS and AT domains have been included in at least one of the engineered enzymes. Thus, there is considerable flexibility in the design of new PKS enzymes with the result that 20 known polyketides can be produced more effectively, and novel polyketides useful as pharmaceuticals or for other purposes can be made.

By appropriate application of recombinant DNA technology, a wide variety of polyketides can be prepared in a variety of different host cells provided one has access to nucleic acid compounds that encode PKS proteins and polyketide modification enzymes. 25 The present invention helps meet the need for such nucleic acid compounds by providing recombinant vectors that encode the FK-520 PKS enzyme and various FK-520 modification enzymes. Moreover, while the FK-506 and FK-520 polyketides have many useful activities, there remains a need for compounds with similar useful activities but with better pharmacokinetic profile and metabolism and fewer side-effects. The present 30 invention helps meet the need for such compounds as well.

Summary of the Invention

In one embodiment, the present invention provides recombinant DNA vectors that encode all or part of the FK-520 PKS enzyme. Illustrative vectors of the invention 35 include cosmid pKOS034-120, pKOS034-124, pKOS065-C31, pKOS065-C3,

WO 00/20601

pKOS065-M27, and pKOS065-M21. The invention also provides nucleic acid compounds that encode the various domains of the FK-520 PKS, i.e., the KS, AT, ACP, KR, DH, and ER domains. These compounds can be readily used, alone or in combination with nucleic acids encoding other FK-520 or non-FK-520 PKS domains, as 5 intermediates in the construction of recombinant vectors that encode all or part of PKS enzymes that make novel polyketides.

The invention also provides isolated nucleic acids that encode all or part of one or more modules of the FK-520 PKS, each module comprising a ketosynthase activity, an acyl transferase activity, and an acyl carrier protein activity. The invention provides an 10 isolated nucleic acid that encodes one or more open reading frames of FK-520 PKS genes, said open reading frames comprising coding sequences for a CoA ligase activity, an NRPS activity, or two or more extender modules. The invention also provides recombinant expression vectors containing these nucleic acids.

In another embodiment, the invention provides isolated nucleic acids that encode 15 all or a part of a PKS that contains at least one module in which at least one of the domains in the module is a domain from a non-FK-520 PKS and at least one domain is from the FK-520 PKS. The non-FK-520 PKS domain or module originates from the rapamycin PKS, the FK-506 PKS, DEBS, or another PKS. The invention also provides recombinant expression vectors containing these nucleic acids.

20 In another embodiment, the invention provides a method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector that encodes at least one module of a PKS, said module comprising at least one FK-520 PKS domain, and culturing said host cell under conditions such that said PKS is produced and catalyzes synthesis of said polyketide. In one aspect, the method is practiced with a *Streptomyces* host cell. In another aspect, the polyketide produced is 25 FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-506 or rapamycin.

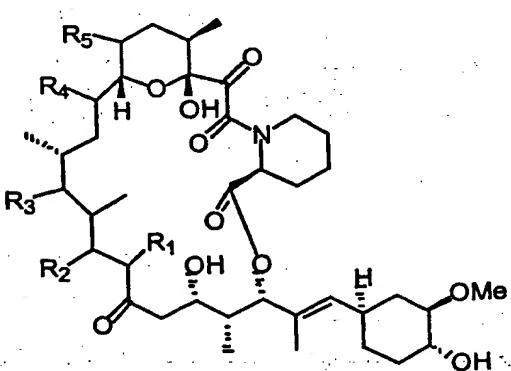
30 In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of ethylmalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require ethylmalonyl CoA for biosynthesis. The invention also provides recombinant nucleic acids that encode AT domains specific for ethylmalonyl CoA. Thus, the compounds of the invention can be

used to produce polyketides requiring ethylmalonyl CoA in host cells that otherwise are unable to produce such polyketides.

In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require 2-hydroxymalonyl CoA for biosynthesis. The invention also provides recombinant nucleic acids that encode AT domains specific for 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring 2-hydroxymalonyl CoA or 2-methoxymalonyl CoA in host cells that are otherwise unable to produce such polyketides.

In another embodiment, the invention provides a compound related in structure to FK-520 or FK-506 that is useful in the treatment of a medical condition. These compounds include compounds in which the C-13 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. Such compounds are less susceptible to the main *in vivo* pathway of degradation for FK-520 and FK-506 and related compounds and thus exhibit an improved pharmacokinetic profile. The compounds of the invention also include compounds in which the C-15 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. The compounds of the invention also include the above compounds further modified by chemical methodology to produce derivatives such as, but not limited to, the C-18 hydroxyl derivatives, which have potent neurotrophin but not immunosuppression activities.

Thus, the invention provides polyketides having the structure:



25

wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen

or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506. The invention provides these compounds in purified form and in pharmaceutical compositions.

5 In another embodiment, the invention provides a method for treating a medical condition by administering a pharmaceutically efficacious dose of a compound of the invention. The compounds of the invention may be administered to achieve immunosuppression or to stimulate nerve growth and regeneration.

These and other embodiments and aspects of the invention will be more fully
10 understood after consideration of the attached Drawings and their brief description
below, together with the detailed description, examples, and claims that follow.

Brief Description of the Drawings

Figure 1 shows a diagram of the FK-520 biosynthetic gene cluster. The top line provides a scale in kilobase pairs (kb). The second line shows a restriction map with selected restriction enzyme recognition sequences indicated. K is *KpnI*; X is *XhoI*, S is *SacI*; P is *PstI*; and E is *EcoRI*. The third line indicates the position of FK-520 PKS and related genes. Genes are abbreviated with a one letter designation, i.e., C is *fkbC*. Immediately under the third line are numbered segments showing where the loading
20 module (L) and ten different extender modules (numbered 1 - 10) are encoded on the various genes shown. At the bottom of the Figure, the DNA inserts of various cosmids of the invention (i.e., 34-124 is cosmid pKOS034-124) are shown in alignment with the FK-520 biosynthetic gene cluster.

Figure 2 shows the loading module (load), the ten extender modules, and the peptide synthetase domain of the FK-520 PKS, together with, on the top line, the genes that encode the various domains and modules. Also shown are the various intermediates in FK-520 biosynthesis, as well as the structure of FK-520, with carbons 13, 15, 21, and 31 numbered. The various domains of each module and subdomains of the loading module are also shown. The darkened circles showing the DH domains in modules 2, 3, and 4 indicate that the dehydratase domain is not functional as a dehydratase; this domain may affect the stereochemistry at the corresponding position in the polyketide. The substituents on the FK-520 structure that result from the action of non-PKS enzymes are also indicated by arrows, together with the types of enzymes or the genes that code for the enzymes that mediate the action. Although the methyltransferase is shown acting at the C-13 and C-15 hydroxyl groups after release of the polyketide from the PKS, the

methyltransferase may act on the 2-hydroxymalonyl substrate prior to or contemporaneously with its incorporation during polyketide synthesis.

Figure 3 shows a close-up view of the left end of the FK-520 gene cluster, which contains at least ten additional genes. The ethyl side chain on carbon 21 of FK-520 (Figure 2) is derived from an ethylmalonyl CoA extender unit that is incorporated by an ethylmalonyl specific AT domain in extender module 4 of the PKS. At least four of the genes in this region code for enzymes involved in ethylmalonyl biosynthesis. The polyhydroxybutyrate depolymerase is involved in maintaining hydroxybutyryl-CoA pools during FK-520 production. Polyhydroxybutyrate accumulates during vegetative growth and disappears during stationary phase in other *Streptomyces* (Ranade and Vining, 1993, *Can. J. Microbiol.* 39:377). Open reading frames with unknown function are indicated with a question mark.

Figure 4 shows a biosynthetic pathway for the biosynthesis of ethylmalonyl CoA from acetoacetyl CoA consistent with the function assigned to four of the genes in the FK-520 gene cluster shown in Figure 3.

Figure 5 shows a close-up view of the right-end of the FK-520 PKS gene cluster (and of the sequences on cosmid pKOS065-C31). The genes shown include *fkbD*, *fkbM* (a methyl transferase that methylates the hydroxyl group on C-31 of FK-520), *fkbN* (a homolog of a gene described as a regulator of cholesterol oxidase and that is believed to be a transcriptional activator), *fkbQ* (a type II thioesterase, which can increase polyketide production levels), and *fkbS* (a crotonyl-CoA reductase involved in the biosynthesis of ethylmalonyl CoA).

Figure 6 shows the proposed degradative pathway for tacrolimus (FK-506) metabolism.

Figure 7 shows a schematic process for the construction of recombinant PKS genes of the invention that encode PKS enzymes that produce 13-desmethoxy FK-506 and FK-520 polyketides of the invention, as described in Example 4, below.

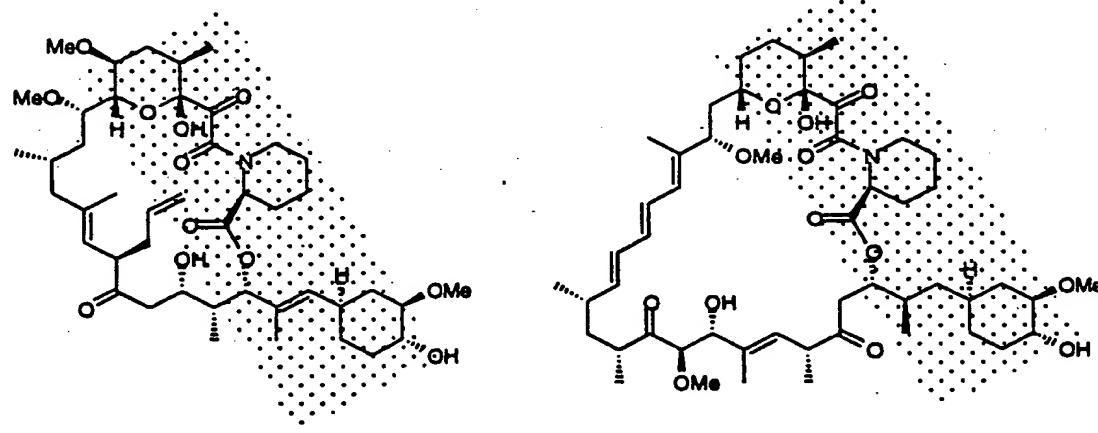
Figure 8, in Parts A and B, shows certain compounds of the invention preferred for dermal application in Part A and a synthetic route for making those compounds in Part B.

Detailed Description of the Invention

Given the valuable pharmaceutical properties of polyketides, there is a need for methods and reagents for producing large quantities of polyketides, as well as for producing related compounds not found in nature. The present invention provides such

methods and reagents, with particular application to methods and reagents for producing the polyketides known as FK-520, also known as ascomycin or L-683,590 (see Holt *et al.*, 1993, *JACS* 115:9925), and FK-506, also known as tacrolimus. Tacrolimus is a macrolide immunosuppressant used to prevent or treat rejection of transplanted heart, kidney, liver, lung, pancreas, and small bowel allografts. The drug is also useful for the prevention and treatment of graft-versus-host disease in patients receiving bone marrow transplants, and for the treatment of severe, refractory uveitis. There have been additional reports of the unapproved use of tacrolimus for other conditions, including alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, and scleroderma. The invention provides methods and reagents for making novel polyketides related in structure to FK-520 and FK-506, and structurally related polyketides such as rapamycin.

The FK-506 and rapamycin polyketides are potent immunosuppressants, with chemical structures shown below.



15

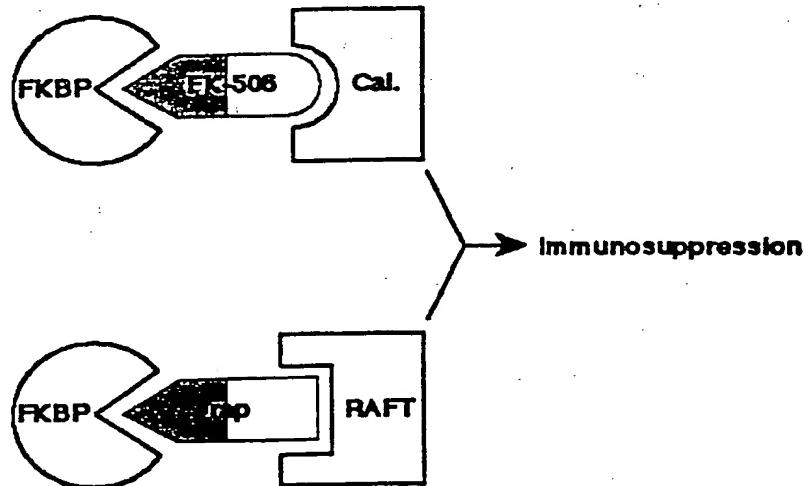
FK-506

Rapamycin

FK-520 differs from FK-506 in that it lacks the allyl group at C-21 of FK-506, having instead an ethyl group at that position, and has similar activity to FK-506, albeit reduced immunosuppressive activity.

These compounds act through initial formation of an intermediate complex with protein "imunophilins" known as FKBP (FK-506 binding proteins), including FKBP-12. Immunophilins are a class of cytosolic proteins that form complexes with molecules such as FK-506, FK-520, and rapamycin that in turn serve as ligands for other cellular targets involved in signal transduction. Binding of FK-506, FK-520, and rapamycin to FKBP occurs through the structurally similar segments of the polyketide molecules, known as the "FKBP-binding domain" (as generally but not precisely indicated by the

stippled regions in the structures above). The FK-506-FKBP complex then binds calcineurin, while the rapamycin-FKBP complex binds to a protein known as RAFT-1. Binding of the FKBP-polyketide complex to these second proteins occurs through the dissimilar regions of the drugs known as the "effector" domains.



5

The three component FKBP-polyketide-effector complex is required for signal transduction and subsequent immunosuppressive activity of FK-506, FK-520, and rapamycin. Modifications in the effector domains of FK-506, FK-520, and rapamycin that destroy binding to the effector proteins (calcineurin or RAFT) lead to loss of 10 immunosuppressive activity, even though FKBP binding is unaffected. Further, such analogs antagonize the immunosuppressive effects of the parent polyketides, because they compete for FKBP. Such non-immunosuppressive analogs also show reduced toxicity (see Dumont *et al.*, 1992, *Journal of Experimental Medicine* 176, 751-760), indicating that much of the toxicity of these drugs is not linked to FKBP binding.

15 In addition to immunosuppressive activity, FK-520, FK-506, and rapamycin have neurotrophic activity. In the central nervous system and in peripheral nerves, immunophilins are referred to as "neuroimmunophilins". The neuroimmunophilin FKBP is markedly enriched in the central nervous system and in peripheral nerves. Molecules that bind to the neuroimmunophilin FKBP, such as FK-506 and FK-520, have the 20 remarkable effect of stimulating nerve growth. *In vitro*, they act as neurotrophins, i.e., they promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. See Gold *et al.*, Jun. 1999, *J. Pharm. Exp. Ther.* 289(3): 1202-1210; Lyons *et al.*, 1994, *Proc. National 25 Academy of Science* 91: 3191-3195; Gold *et al.*, 1995, *Journal of Neuroscience* 15:

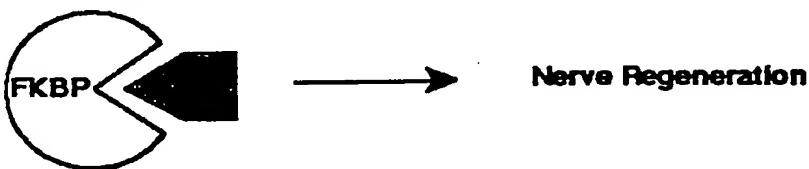
7509-7516; and Steiner *et al.*, 1997, *Proc. National Academy of Science* 94: 2019-2024.

Further, the restored central and peripheral neurons appear to be functional.

- Compared to protein neurotrophic molecules (BDNF, NGF, etc.), the small-molecule neurotrophins such as FK-506, FK-520, and rapamycin have different, and often advantageous, properties. First, whereas protein neurotrophins are difficult to deliver to their intended site of action and may require intra-cranial injection, the small-molecule neurotrophins display excellent bioavailability; they are active when administered subcutaneously and orally. Second, whereas protein neurotrophins show quite specific effects, the small-molecule neurotrophins show rather broad effects.
- 5 Finally, whereas protein neurotrophins often show effects on normal sensory nerves, the small-molecule neurotrophins do not induce aberrant sprouting of normal neuronal processes and seem to affect damaged nerves specifically. Neuroimmunophilin ligands have potential therapeutic utility in a variety of disorders involving nerve degeneration (e.g. multiple sclerosis, Parkinson's disease, Alzheimer's disease, stroke, traumatic
- 10 spinal cord and brain injury, peripheral neuropathies).
- 15

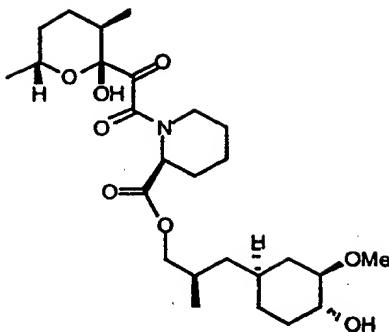
Recent studies have shown that the immunosuppressive and neurite outgrowth activity of FK-506, FK-520, and rapamycin can be separated; the neuroregenerative activity in the absence of immunosuppressive activity is retained by agents which bind to FKBP but not to the effector proteins calcineurin or RAPT. See Steiner *et al.*, 1997,

20 *Nature Medicine* 3: 421-428.



Available structure-activity data show that the important features for neurotrophic activity of rapamycin, FK-520, and FK-506 lie within the common, contiguous segments of the macrolide ring that bind to FKBP. This portion of the molecule is termed the "FKBP binding domain" (see VanDuyne *et al.*, 1993, *Journal of Molecular Biology* 229: 105-124.). Nevertheless, the effector domains of the parent macrolides contribute to conformational rigidity of the binding domain and thus indirectly contribute to FKBP binding.

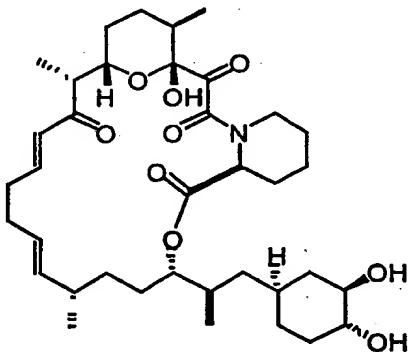
25



"FKBP binding domain"

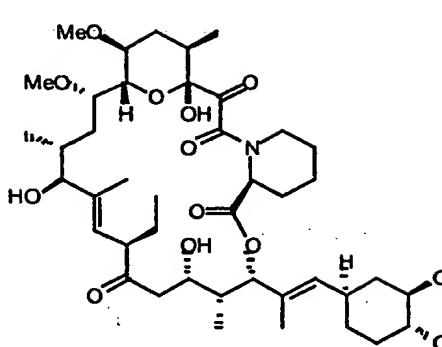
There are a number of other reported analogs of FK-506, FK-520, and rapamycin that bind to FKBP but not the effector protein calcineurin or RAFT. These analogs show effects on nerve regeneration without immunosuppressive effects.

5 Naturally occurring FK-520 and FK-506 analogs include the antascomycins, which are FK-506-like macrolides that lack the functional groups of FK-506 that bind to calcineurin (see Fehr *et al.*, 1996, *The Journal of Antibiotics* 49: 230-233). These molecules bind FKBP as effectively as does FK-506; they antagonize the effects of both FK-506 and rapamycin, yet lack immunosuppressive activity.

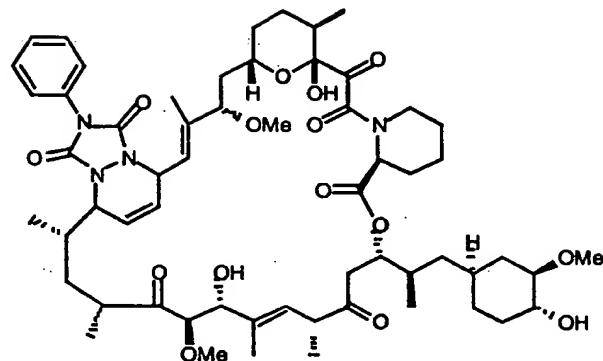


Antascomycin A

10 Other analogs can be produced by chemically modifying FK-506, FK-520, or rapamycin. One approach to obtaining neuroimmunophilin ligands is to destroy the effector binding region of FK-506, FK-520, or rapamycin by chemical modification. While the chemical modifications permitted on the parent compounds are quite limited, 15 some useful chemically modified analogs exist. The FK-520 analog L-685,818 ($ED_{50} = 0.7 \text{ nM}$ for FKBP binding; see Dumont *et al.*, 1992), and the rapamycin analog WAY-124,466 ($IC_{50} = 12.5 \text{ nM}$; see Ocain *et al.*, 1993, *Biochemistry Biophysical Research Communications* 192: 1340-134693) are about as effective as FK-506, FK-520, and rapamycin at promoting neurite outgrowth in sensory neurons (see Steiner *et al.*, 1997).

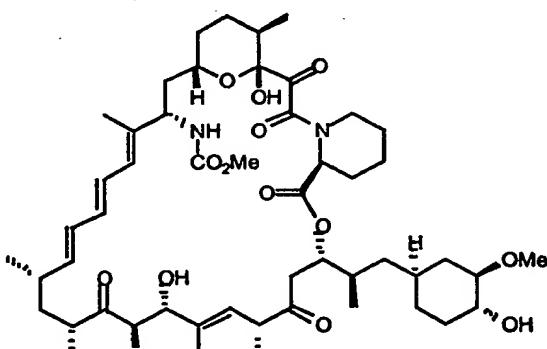


L-685,818



WAY-124,466

One of the few positions of rapamycin that is readily amenable to chemical modification is the allylic 16-methoxy group; this reactive group is readily exchanged by acid-catalyzed nucleophilic substitution. Replacement of the 16-methoxy group of 5 rapamycin with a variety of bulky groups has produced analogs showing selective loss of immunosuppressive activity while retaining FKBP-binding (see Luengo *et al.*, 1995, *Chemistry & Biology* 2: 471-481). One of the best compounds, 1, below, shows complete loss of activity in the splenocyte proliferation assay with only a 10-fold reduction in binding to FKBP.

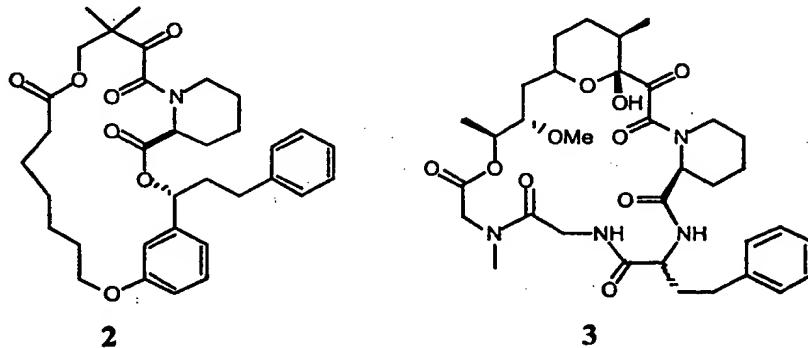


10

1

There are also synthetic analogs of FKBP binding domains. These compounds reflect an approach to obtaining neuroimmunophilin ligands based on "rationally designed" molecules that retain the FKBP-binding region in an appropriate conformation 15 for binding to FKBP, but do not possess the effector binding regions. In one example, the ends of the FKBP binding domain were tethered by hydrocarbon chains (see Holt *et al.*, 1993, *Journal of the American Chemical Society* 115: 9925-9938); the best analog, 2, below, binds to FKBP about as well as FK-506. In a similar approach, the ends of the FKBP binding domain were tethered by a tripeptide to give analog 3, below, which binds

to FKBP about 20-fold poorer than FK-506. These compounds are anticipated to have neuroimmunophilin binding activity.



5 In a primate MPTP model of Parkinson's disease, administration of FKBP ligand
GPI-1046 caused brain cells to regenerate and behavioral measures to improve. MPTP is
a neurotoxin, which, when administered to animals, selectively damages nigral-striatal
dopamine neurons in the brain, mimicking the damage caused by Parkinson's disease.
Whereas, before treatment, animals were unable to use affected limbs, the FKBP ligand
10 restored the ability of animals to feed themselves and gave improvements in measures of
locomotor activity, neurological outcome, and fine motor control. There were also
corresponding increases in regrowth of damaged nerve terminals. These results
demonstrate the utility of FKBP ligands for treatment of diseases of the CNS.

From the above description, two general approaches towards the design of non-
15 immunosuppressant, neuroimmunophilin ligands can be seen. The first involves the
construction of constrained cyclic analogs of FK-506 in which the FKBP binding domain
is fixed in a conformation optimal for binding to FKBP. The advantages of this approach
are that the conformation of the analogs can be accurately modeled and predicted by
computational methods, and the analogs closely resemble parent molecules that have
20 proven pharmacological properties. A disadvantage is that the difficult chemistry limits
the numbers and types of compounds that can be prepared. The second approach
involves the trial and error construction of acyclic analogs of the FKBP binding domain
by conventional medicinal chemistry. The advantages to this approach are that the
chemistry is suitable for production of the numerous compounds needed for such
25 interactive chemistry-bioassay approaches. The disadvantages are that the molecular
types of compounds that have emerged have no known history of appropriate
pharmacological properties, have rather labile ester functional groups, and are too
conformationally mobile to allow accurate prediction of conformational properties.

The present invention provides useful methods and reagents related to the first
30 approach, but with significant advantages. The invention provides recombinant PKS

genes that produce a wide variety of polyketides that cannot otherwise be readily synthesized by chemical methodology alone. Moreover, the present invention provides polyketides that have either or both of the desired immunosuppressive and neurotrophic activities, some of which are produced only by fermentation and others of which are 5 produced by fermentation and chemical modification. Thus, in one aspect, the invention provides compounds that optimally bind to FKBP but do not bind to the effector proteins. The methods and reagents of the invention can be used to prepare numerous constrained cyclic analogs of FK-520 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. Such compounds will show 10 neuroimmunophilin binding (neurotrophic) but not immunosuppressive effects. The invention also allows direct manipulation of FK-520 and related chemical structures *via* genetic engineering of the enzymes involved in the biosynthesis of FK-520 (as well as related compounds, such as FK-506 and rapamycin); similar chemical modifications are simply not possible because of the complexity of the structures. The invention can also 15 be used to introduce "chemical handles" into normally inert positions that permit subsequent chemical modifications.

Several general approaches to achieve the development of novel neuroimmunophilin ligands are facilitated by the methods and reagents of the present invention. One approach is to make "point mutations" of the functional groups of the 20 parent FK-520 structure that bind to the effector molecules to eliminate their binding potential. These types of structural modifications are difficult to perform by chemical modification, but can be readily accomplished with the methods and reagents of the invention.

A second, more extensive approach facilitated by the present invention is to 25 utilize molecular modeling to predict optimal structures *ab initio* that bind to FKBP but not effector molecules. Using the available X-ray crystal structure of FK-520 (or FK-506) bound to FKBP, molecular modeling can be used to predict polyketides that should optimally bind to FKBP but not calcineurin. Various macrolide structures can be generated by linking the ends of the FKBP-binding domain with "all possible" 30 polyketide chains of variable length and substitution patterns that can be prepared by genetic manipulation of the FK-520 or FK-506 PKS gene cluster in accordance with the methods of the invention. The ground state conformations of the virtual library can be determined, and compounds that possess binding domains most likely to bind well to FKBP can be prepared and tested.

Once a compound is identified in accordance with the above approaches, the invention can be used to generate a focused library of analogs around the lead candidate, to "fine tune" the compound for optimal properties. Finally, the genetic engineering methods of the invention can be directed towards producing "chemical handles" that enable medicinal chemists to modify positions of the molecule previously inert to chemical modification. This opens the path to previously prohibited chemical optimization of lead compounds by time-proven approaches.

Moreover, the present invention provides polyketide compounds and the recombinant genes for the PKS enzymes that produce the compounds that have significant advantages over FK-506 and FK-520 and their analogs. The metabolism and pharmacokinetics of tacrolimus has been extensively studied, and FK-520 is believed to be similar in these respects. Absorption of tacrolimus is rapid, variable, and incomplete from the gastrointestinal tract (Harrison's Principles of Internal Medicine, 14th edition, 1998, McGraw Hill, 14, 20, 21, 64-67). The mean bioavailability of the oral dosage form is 27%, (range 5 to 65%). The volume of distribution (V_oD) based on plasma is 5 to 65 L per kg of body weight (L/kg), and is much higher than the V_oD based on whole blood concentrations, the difference reflecting the binding of tacrolimus to red blood cells. Whole blood concentrations may be 12 to 67 times the plasma concentrations. Protein binding is high (75 to 99%), primarily to albumin and alpha₁-acid glycoprotein. The half-life for distribution is 0.9 hour; elimination is biphasic and variable: terminal-11.3 hr (range, 3.5 to 40.5 hours). The time to peak concentration is 0.5 to 4 hours after oral administration.

Tacrolimus is metabolized primarily by cytochrome P450 3A enzymes in the liver and small intestine. The drug is extensively metabolized with less than 1% excreted unchanged in urine. Because hepatic dysfunction decreases clearance of tacrolimus, doses have to be reduced substantially in primary graft non-function, especially in children. In addition, drugs that induce the cytochrome P450 3A enzymes reduce tacrolimus levels, while drugs that inhibit these P450s increase tacrolimus levels. Tacrolimus bioavailability doubles with co-administration of ketoconazole, a drug that inhibits P450 3A. See, Vincent *et al.*, 1992, *In vitro* metabolism of FK-506 in rat, rabbit, and human liver microsomes: Identification of a major metabolite and of cytochrome P450 3A as the major enzymes responsible for its metabolism, *Arch. Biochem. Biophys.* 294: 454-460; Iwasaki *et al.*, 1993, Isolation, identification, and biological activities of oxidative metabolites of FK-506, a potent immunosuppressive macrolide lactone, *Drug Metabolism & Disposition* 21: 971-977; Shiraga *et al.*, 1994, Metabolism of FK-506, a

potent immunosuppressive agent, by cytochrome P450 3A enzymes in rat, dog, and human liver microsomes, *Biochem. Pharmacol.* 47: 727-735; and Iwasaki *et al.*, 1995, Further metabolism of FK-506 (Tacrolimus); Identification and biological activities of the metabolites oxidized at multiple sites of FK-506, *Drug Metabolism & Disposition* 23:

- 5 28-34. The cytochrome P450 3A subfamily of isozymes has been implicated as important in this degradative process.

Structures of the eight isolated metabolites formed by liver microsomes are shown in Figure 6. Four metabolites of FK-506 involve demethylation of the oxygens on carbons 13, 15, and 31, and hydroxylation of carbon 12. The 13-demethylated (hydroxy) 10 compounds undergo cyclizations of the 13-hydroxy at C-10 to give M-I, M-VI and M-VII, and the 12-hydroxy metabolite at C-10 to give I. Another four metabolites formed by oxidation of the four metabolites mentioned above were isolated by liver microsomes from dexamethasone treated rats. Three of these are metabolites doubly demethylated at the methoxy groups on carbons 15 and 31 (M-V), 13 and 31 (M-VI), and 13 and 15 (M-VII). The fourth, M-VIII, was the metabolite produced after demethylation of the 31-methoxy group, followed by formation of a fused ring system by further oxidation. Among the eight metabolites, M-II has immunosuppressive activity comparable to that of FK-506, whereas the other metabolites exhibit weak or negligible activities. Importantly, the major metabolite of human, dog, and rat liver microsomes is the 13-demethylated and cyclized FK-506 (M-I).

Thus, the major metabolism of FK-506 proceeds via 13-demethylation followed by cyclization to the inactive M-I, this representing about 90% of the metabolic products after a 10 minute incubation with liver microsomes. Analogs of tacrolimus that do not possess a C-13 methoxy group would not be susceptible to the first and most important 25 biotransformation in the destructive metabolism of tacrolimus (i.e. cyclization of 13-hydroxy to C-10). Thus, a 13-desmethoxy analog of FK-506 should have a longer half-life in the body than does FK-506. The C-13 methoxy group is believed not to be required for binding to FKBP or calcineurin. The C-13 methoxy is not present on the identical position of rapamycin, which binds to FKBP with equipotent affinity as 30 tacrolimus. Also, analysis of the 3-dimensional structure of the FKBP-tacrolimus-calcineurin complex shows that the C-13 methoxy has no interaction with FKBP and only a minor interaction with calcineurin. The present invention provides C-13-desmethoxy analogs of FK-506 and FK-520, as well as the recombinant genes that encode the PKS enzymes that catalyze their synthesis and host cells that produce the 35 compounds.

These compounds exhibit, relative to their naturally occurring counterparts, prolonged immunosuppressive action *in vivo*, thereby allowing a lower dosage and/or reduced frequency of administration. Dosing is more predictable, because the variability in FK-506 dosage is largely due to variation of metabolism rate. FK-506 levels in blood
5 can vary widely depending on interactions with drugs that induce or inhibit cytochrome P450 3A (summarized in USP Drug Information for the Health Care Professional). Of particular importance are the numerous drugs that inhibit or compete for CYP 3A, because they increase FK-506 blood levels and lead to toxicity (Prograf package insert, Fujisawa□US, Rev 4/97, Rec 6/97). Also important are the drugs that induce P450 3A
10 (e.g. Dexamethasone), because they decrease FK-506 blood levels and reduce efficacy. Because the major site of CYP 3A action on FK-506 is removed in the analogs provided by the present invention, those analogs are not as susceptible to drug interactions as the naturally occurring compounds.

Hyperglycemia, nephrotoxicity, and neurotoxicity are the most significant
15 adverse effects resulting from the use of FK-506 and are believed to be similar for FK-520. Because these effects appear to occur primarily by the same mechanism as the immunosuppressive action (i.e. FKBP-calcineurin interaction), the intrinsic toxicity of the desmethoxy analogs may be similar to FK-506. However, toxicity of FK-506 is dose related and correlates with high blood levels of the drug (Prograf package insert,
20 Fujisawa□US, Rev 4/97, Rec 6/97). Because the levels of the compounds provided by the present invention should be more controllable, the incidence of toxicity should be significantly decreased with the 13-desmethoxy analogs. Some reports show that certain FK-506 metabolites are more toxic than FK-506 itself, and this provides an additional reason to expect that a CYP 3A resistant analog can have lower toxicity and a higher
25 therapeutic index.

Thus, the present invention provides novel compounds related in structure to FK-506 and FK-520 but with improved properties. The invention also provides methods for making these compounds by fermentation of recombinant host cells, as well as the recombinant host cells, the recombinant vectors in those host cells, and the recombinant
30 proteins encoded by those vectors. The present invention also provides other valuable materials useful in the construction of these recombinant vectors that have many other important applications as well. In particular, the present invention provides the FK-520 PKS genes, as well as certain genes involved in the biosynthesis of FK-520 in recombinant form.

FK-520 is produced at relatively low levels in the naturally occurring cells, *Streptomyces hygroscopicus* var. *ascomyceticus*, in which it was first identified. Thus, another benefit provided by the recombinant FK-520 PKS and related genes of the present invention is the ability to produce FK-520 in greater quantities in the 5 recombinant host cells provided by the invention. The invention also provides methods for making novel FK-520 analogs, in addition to the desmethoxy analogs described above, and derivatives in recombinant host cells of any origin.

The biosynthesis of FK-520 involves the action of several enzymes. The FK-520 PKS enzyme, which is composed of the *fkbA*, *fkbB*, *fkbC*, and *fkbP* gene products, 10 synthesizes the core structure of the molecule. There is also a hydroxylation at C-9 mediated by the P450 hydroxylase that is the *fkbD* gene product and that is oxidized by the *fkbO* gene product to result in the formation of a keto group at C-9. There is also a methylation at C-31 that is mediated by an O-methyltransferase that is the *fkbM* gene product. There are also methylations at the C-13 and C-15 positions by a 15 methyltransferase believed to be encoded by the *fkbG* gene; this methyltransferase may act on the hydroxymalonyl CoA substrates prior to binding of the substrate to the AT domains of the PKS during polyketide synthesis. The present invention provides the genes encoding these enzymes in recombinant form. The invention also provides the genes encoding the enzymes involved in ethylmalonyl CoA and 2-hydroxymalonyl CoA 20 biosynthesis in recombinant form. Moreover, the invention provides *Streptomyces hygroscopicus* var. *ascomyceticus* recombinant host cells lacking one or more of these genes that are useful in the production of useful compounds.

The cells are useful in production in a variety of ways. First, certain cells make a useful FK-520-related compound merely as a result of inactivation of one or more of the 25 FK-520 biosynthesis genes. Thus, by inactivating the C-31 O-methyltransferase gene in *Streptomyces hygroscopicus* var. *ascomyceticus*, one creates a host cell that makes a desmethyl (at C-31) derivative of FK-520. Second, other cells of the invention are unable to make FK-520 or FK-520 related compounds due to an inactivation of one or more of the PKS genes. These cells are useful in the production of other polyketides produced by 30 PKS enzymes that are encoded on recombinant expression vectors and introduced into the host cell.

Moreover, if only one PKS gene is inactivated, the ability to produce FK-520 or an FK-520 derivative compound is restored by introduction of a recombinant expression vector that contains the functional gene in a modified or unmodified form. The 35 introduced gene produces a gene product that, together with the other endogenous and

functional gene products, produces the desired compound. This methodology enables one to produce FK-520 derivative compounds without requiring that all of the genes for the PKS enzyme be present on one or more expression vectors. Additional applications and benefits of such cells and methodology will be readily apparent to those of skill in the art after consideration of how the recombinant genes were isolated and employed in the construction of the compounds of the invention.

The FK-520 biosynthetic genes were isolated by the following procedure. Genomic DNA was isolated from *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) using the lysozyme/proteinase K protocol described in Genetic 10 Manipulation of *Streptomyces* - A Laboratory Manual (Hopwood *et al.*, 1986). The average size of the DNA was estimated to be between 80 - 120 kb by electrophoresis on 0.3% agarose gels. A library was constructed in the SuperCos™ vector according to the manufacturer's instructions and with the reagents provided in the commercially available kit (Stratagene). Briefly, 100 µg of genomic DNA was partially digested with 4 units of 15 *Sau3A* I for 20 min. in a reaction volume of 1 mL, and the fragments were dephosphorylated and ligated to SuperCos vector arms. The ligated DNA was packaged and used to infect log-stage XL1-BlueMR cells. A library of about 10,000 independent cosmid clones was obtained.

Based on recently published sequence from the FK-506 cluster (Motamedi and 20 Shafiee, 1998, *Eur. J. Biochem.* 256: 528), a probe for the *fkbO* gene was isolated from ATCC 14891 using PCR with degenerate primers. With this probe, a cosmid designated pKOS034-124 was isolated from the library. With probes made from the ends of cosmid pKOS034-124, an additional cosmid designated pKOS034-120 was isolated. These 25 cosmids (pKOS034-124 and pKOS034-120) were shown to contain DNA inserts that overlap with one another. Initial sequence data from these two cosmids generated sequences similar to sequences from the FK-506 and rapamycin clusters, indicating that the inserts were from the FK-520 PKS gene cluster. Two *EcoRI* fragments were subcloned from cosmids pKOS034-124 and pKOS034-120. These subclones were used 30 to prepare shotgun libraries by partial digestion with *Sau3AI*, gel purification of fragments between 1.5 kb and 3 kb in size, and ligation into the pLitmus28 vector (New England Biolabs). These libraries were sequenced using dye terminators on a Beckmann CEQ2000 capillary electrophoresis sequencer, according to the manufacturer's protocols.

To obtain cosmids containing sequence on the left and right sides of the 35 sequenced region described above, a new cosmid library of ATCC 14891 DNA was prepared essentially as described above. This new library was screened with a new *fkbM*

probe isolated using DNA from ATCC 14891. A probe representing the *fkbP* gene at the end of cosmid pKOS034-124 was also used. Several additional cosmids to the right of the previously sequenced region were identified. Cosmids pKOS065-C31 and pKOS065-C3 were identified and then mapped with restriction enzymes. Initial sequences from 5 these cosmids were consistent with the expected organization of the cluster in this region. More extensive sequencing showed that both cosmids contained in addition to the desired sequences, other sequences not contiguous to the desired sequences on the host cell chromosomal DNA. Probing of additional cosmid libraries identified two additional cosmids, pKOS065-M27 and pKOS065-M21, that contained the desired sequences in a 10 contiguous segment of chromosomal DNA. Cosmids pKOS034-124, pKOS034-120, pKOS065-M27, and pKOS065-M21 have been deposited with the American Type Culture Collection, Manassas, VA, USA. The complete nucleotide sequence of the coding sequences of the genes that encode the proteins of the FK-520 PKS are shown 15 below but can also be determined from the cosmids of the invention deposited with the ATCC using standard methodology.

Referring to Figures 1 and 3, the FK-520 PKS gene cluster is composed of four open reading frames designated *fkbB*, *fkbC*, *fkbA*, and *fkbP*. The *fkbB* open reading frame encodes the loading module and the first four extender modules of the PKS. The *fkbC* open reading frame encodes extender modules five and six of the PKS. The *fkbA* open 20 reading frame encodes extender modules seven, eight, nine, and ten of the PKS. The *fkbP* open reading frame encodes the NRPS of the PKS. Each of these genes can be isolated from the cosmids of the invention described above. The DNA sequences of these genes are provided below preceded by the following table identifying the start and stop codons of the open reading frames of each gene and the modules and domains contained 25 therein.

	<u>Nucleotides</u>	<u>Gene or Domain</u>
	complement (412 - 1836)	<i>fkbW</i>
	complement (2020 - 3579)	<i>fkbV</i>
30	complement (3969 - 4496)	<i>fkbR2</i>
	complement (4595 - 5488)	<i>fkbR1</i>
	5601 - 6818	<i>fkbE</i>
	6808 - 8052	<i>fkbF</i>
	8156 - 8824	<i>fkbG</i>
35	complement (9122 - 9883)	<i>fkbH</i>
	complement (9894 - 10994)	<i>fkbI</i>
	complement (10987 - 11247)	<i>fkbJ</i>
	complement (11244 - 12092)	<i>fkbK</i>
	complement (12113 - 13150)	<i>fkbL</i>
40	complement (13212 - 23988)	<i>fkbC</i>

	complement (23992 - 46573)	<i>fkbB</i>
	46754 - 47788	<i>fkbO</i>
	47785 - 52272	<i>fkbP</i>
	52275 - 71465	<i>fkbA</i>
5	71462 - 72628	<i>fkbD</i>
	72625 - 73407	<i>fkbM</i>
	complement (73460 - 76202)	<i>fkbN</i>
	complement (76336 - 77080)	<i>fkbQ</i>
	complement (77076 - 77535)	<i>fkbS</i>
10	complement (44974 - 46573)	CoA ligase of loading domain
	complement (43777 - 44629)	ER of loading domain
	complement (43144 - 43660)	ACP of loading domain
	complement (41842 - 43093)	KS of extender module 1 (KS1)
	complement (40609 - 41842)	AT1
15	complement (39442 - 40609)	DH1
	complement (38677 - 39307)	KR1
	complement (38371 - 38581)	ACP1
	complement (37145 - 38296)	KS2
	complement (35749 - 37144)	AT2
20	complement (34606 - 35749)	DH2 (inactive)
	complement (33823 - 34480)	KR2
	complement (33505 - 33715)	ACP2
	complement (32185 - 33439)	KS3
	complement (31018 - 32185)	AT3
25	complement (29869 - 31018)	DH3 (inactive)
	complement (29092 - 29740)	KR3
	complement (28750 - 28960)	ACP3
	complement (27430 - 28684)	KS4
	complement (26146 - 27430)	AT4
30	complement (24997 - 26146)	DH4 (inactive)
	complement (24163 - 24373)	ACP4
	complement (22653 - 23892)	KS5
	complement (21420 - 22653)	AT5
	complement (20241 - 21420)	DHS
35	complement (19464 - 20097)	KR5
	complement (19116 - 19326)	ACP5
	complement (17820 - 19053)	KS6
	complement (16587 - 17820)	AT6
	complement (15438 - 16587)	DH6
40	complement (14517 - 15294)	ER6
	complement (13761 - 14394)	KR6
	complement (13452 - 13662)	ACP6
	52362 - 53576	KS7
	53577 - 54716	AT7
45	54717 - 55871	DH7
	56019 - 56819	ER7
	56943 - 57575	KR7
	57710 - 57920	ACP7
	57990 - 59243	KS8
50	59244 - 60398	AT8
	60399 - 61412	DH8 (inactive)
	61548 - 62180	KR8

	62328 - 62537	ACP8
	62598 - 63854	KS9
	63855 - 65084	AT9
	65085 - 66254	DH9
5	66399 - 67175	ER9
	67299 - 67931	KR9
	68094 - 68303	ACP9
	68397 - 69653	KS10
	69654 - 70985	AT10
10	71064 - 71273	ACP10

	1	GATCTCAGGC ATGAAGTCCT CCAGGCGAGG CGCCGAGGTG GTGAACACCT CGCCGCTGCT
	61	TGTACGGACC ACTTCAGTCA GCGGCGATTG CGGAACCAAAG TCATCCGAA TAAAGGGCGG
15	121	TTACAAGATC CTCACATTGC GCGACCGCCA GCATACGCTG AGTTGCCTCA GAGGCAAACC
	181	GAAAGGGCGC GGGCGGTCCG CACCAGGGCG GAGTACGCGA CGAGAGTGGC GCACCCGCGC
	241	ACCGTCACCT CTCTCCCCG CGGGCGGGAT GCCCCGGCTG ACACGGTTGG GCTCTCCTCG
	301	ACGCTGAACA CCCGCGCGGT GTGGCGTCGG GGACACCGCC TGGCATCGC CGGGTGACGG
	361	TACGGGGAGG GCGTACGGCG GCCGTGGCTC GTGCTCACGG CCGCCGGGCG GTCATCCGTC
20	421	GAGACGGCAC TCGGCGAGCA GGGACGCCCTG GTCGGCACCT GCGGGCCCGA CGACCGTGTG
	481	GTTCGCGGGC GGGCGGTGGC CGGTGGTGAN CCAGCTCTCC AGGGCGGTGA AGGCTGAGCG
	541	GTGACACGGC AGCAAAGGCC GGAGTCGGTC GGGGAAGGTG TCGACGAGGG CGTCGGTGTG
	601	CGTGCCGTCC TCGATGCGGT AGTAGCGGTA CCGGCCGCCA GGCCGCTGCC GGACATACGC
	661	GCGTACACGT CGGAGCCC GGCGCAGGCA GCAGCACGTC GAGAGTGCCT GGATGGTGT
25	721	CAGCGGTTG CCGATACGAC CGGTCAACGC GATGCGTTCC ACGGCCGCGT GGACGCJGGA
	781	GGAGCGGGTG GCGTAGTCGT AGTCGGCATC GCAGCCCCGG ACCGTCCCCCG GGGCGCAATA
	841	CGGTGTGCCG GCTTCCTTCT CCCCATCGAA GCGGGGGTCG AACTCCTCGC GGTAGACGCG
	901	CTGCGTCAGA TCCCAGTAGA CCTCGTGGTG CTACGGCCAC AAGAACTCGG AGTCGGCCGG
	961	GAACCCGGCG CGGAGCAGCG CCTCGCCGCG CTGGCCGGCT GCGGGGCCGC CTGCCGCGTA
30	1021	GGTGGGGTAG TCGCGCAGGG CGGGCGGAG GAAAGGTGAAG AGGTTGGGAC CCTCCGCGC
	1081	CCACAGGGTG CCTTCCCAGT CGACTCCTCC GTCGTACAGC TCGGGATGGT TCTCCAGCTG
	1141	CCAGCGCACG AGGTAGCCGC CGTTGGACAT CCCGGTGACC AGGGTGCCT CGAGCGGCCG
	1201	GTGGTAGGCCG TGGGGCACCG ACCGGCGGGC GGGCGGGTC AGCTGGGTGA GGCGGGTGT
	1261	CCACTCGGCCG ACGGCGTCGC CGGGCGGGGA GCCATCACGG TAGAACGCGG GGCGGGTGT
35	1321	GCCCTTGTCTG GTGGCGCGT AGGGCGTAACC GCGGGCGAGC ACCCAGTCGG CGATGGCCCG
	1381	GTCGTTGGCG TACTGCTCGC GGTTACCGGG GGTGCCGGCC ACGACCAGGC CACCGTTCCA
	1441	GCGGTGGGC AGCCGGATGA CGAACTGGGC GTCGTGGTTC CACCCGTGGT TGGTGTGTT
	1501	GGTGGAGGTG TCGGGGAAGT AGCCGTCGAT CTGGATCCCG GGCACCTCCGG TGGGAGTGGC
	1561	CAGGTTCTTG GCGTCAGCC CTGCCAGTC CGCCGGGTG GTGTGGCCGG TGGCCGCCGT
40	1621	TCCCGCCGTG GTCAGCTCGT CCAGGCAGTC GGCCTGCTGA CGTGCCGCC CGGGGACACG
	1681	CAGCTGGGAC AGACGGGCGC AGTACCGTC CGGGGCATCG GGAGCAGGCC GGCGCGTGGC
	1741	CGGTGAGGGG AGCAGGACGG CGACTCGGCC CAGGGTGAGA GCGCCGAGGC CGGTGCGTCT
	1801	TCTCGGGGCC CGTCCGACAC CGAGGGGCAG AACCATGGAG AGCCTCCAGA CGTGCCTGATG
	1861	GATGACGGAC TGGAGGCTAG GTCCGCACG GTGGAGACGA ACATGGGTGC GCCCCTCATG
	1921	ACTGAGGCC CTCAGAGGTG GGCCGCCGCC ATGACGGGCG CGGGACCGCG GCGCTCCGG
45	1981	GGCGGTGCCG CGGGCCGCCA CGGGTCCGG GTCCCCGGGT CAGGGACAGG TGCGTTCGC
	2041	GACGGTGAAG TAGCCGGTCG GCGACTCTTT CAAGGTGGTC GTGACGAAGG TGTGACAG
	2101	GCCCATGTTG TGGCCGGAGC CCTTGGCTA GGTGTAAACCG GCGCTCGTCG TGGCGCGGCC
	2161	CGCCTGGACG TGAGCGTAGT TGCCGGCGGT CCAGCAGACG GCCGTGGCAC CGTCGCTCTG
	2221	CGCGGTGACC GCGCCCGAGA CGGGTCCGGC CTTGCCGTCC GCGTCCCGGG CGCGACCGC
50	2281	GTAGGGTGTGC GATGTGCCCG CCCTCAGGCC GGTGTCCGTG TACGACGTCG TGGCGGACGT
	2341	GGTGATCTGG GCACCGTCGC GGTGGACGGC GTAGTCGGTG GCGCCGTCGA CGGGTTCCA
	2401	GGTCAGGCTG ATGGTGGTGT CGGTGGCGCC GGTGGCGGCC AGGCCGGACG GAGCGGGCAG
	2461	CGAACCGGGG TCGGAGGCCG ATCCGCTCAG GCGAAGAAC TCGTGATCC AGTAGCTGA
55	2521	ACAGATCGAG TCCAGGAAGT AGGCCGCC GGTGCTGCCG CACTGCTGTG CTCCGGTGC
	2581	GGGATCGACC GGGGTGCCGT GCCCCATGCC CGGCACCCGG TTCACCTCCA CGGCCACCGA
	2641	TCCGTCCGCG GCCAGGTACT CCTCGTGCCG GGTGGAGTTC GGGCGATCA CCGAGGTACG
	2701	GTCCGGCGTC TGGGACACGC CGTGCACAGC GGTCCACTGG TCGCGCAACT CGTCGGCGTT
	2761	GCGCGCGCG ACGGTGGTGT CCTTGTGCC GGTGCCAGATG GCCACGCGCG GCCACGGGCC
	2821	CGACACAGAG GGGTAGCCGT CACGGACCG CGCGCCAC TGGTCCGCGG TCAGGTGGT
60	2881	CCCGGGGTTC ATGCACAGGT ACGCCGTGCT GACGTGGTG GCACAGCCGA AGGGCAGGCC
	2941	GGCGACGACC GCGCCGCCCT GGAAGACGTC CGGATAGGTG GCGAGCATCA CCGACGTCAT

	3001	GGCACCGCCG	GCGGACAGCC	CGGTGATGTA	GGTGCCTGG	GGGTCGCCGC	CGTAGGCCGA
	3061	GACGGTGTGA	GCGGCCATCT	GCGGATCGA	CGCGCCTTCG	CCCTGGCCCC	TGCGGTTGTC
5	3121	GCTGCTCTGG	AACCAGTTGA	AGCACCTGTT	CGCGTTGTT	GACGACGTGG	TCTCGCGAA
	3181	CACGAGCAGG	AAGCCATAGC	GGTCCCGCAA	TGAGAGCAGG	CCGGAGTTGT	CGGCGTAGCC
	3241	CTGGGCGTCC	TGGGTGCAAC	CGTGCAGGGC	GAACACCACC	GCCGGCTCCG	CGGGCAGGGA
	3301	CGCGGGCCGG	TAGACGTACA	TGTTCAGCCG	GCCC GGTT	GTGCCGAAGT	CGCGCACCTC
	3361	GGTCAGGTCC	GCCTTGTCA	GACCGGGCTT	GGCCAGGCC	GCCGCGGCGT	GGCCGTCGG
	3421	CGCGGGCCGG	AGCAGGGCCG	CTCCGAGTAC	GAGGGCCACG	ACGGCCACGA	GACGGG" GAG
10	3481	CACCCCCCGC	CGTCCCGGAC	GCGACAAACGA	CCCAGCCG	GGCGAGGAGG	AGAGGGGGAA
	3541	CAGCGGGGTG	AGGATTCCCC	GGAACCGCGG	CGGCTGCATG	CGGCTCCCT	CGATGTCGTG
	3601	GGGGGGACAC	GGAGGGCTCC	CTGACGTG	TCAGTGGGAG	CGCCCCGGTG	CCCGGCACCG
	3661	TAGGGGTGGT	TCAACCGCA	ACGGTATGGC	CCGGAGCAC	ACACCCCGCA	CCGCGCGATG
	3721	TGCGCCCGGA	CGGATTGTGT	CGCCTGCGG	AATCTGATAC	CCGGACGCGA	CGAACGCC
15	3781	ACCCGACACG	GGTAGGGCGT	CATGGTGTCC	GACTCGCCG	GTCCGCCCTG	CCTGCCCTGG
	3841	ACGGACCGGG	CGTCGGCGGA	CGGGCGCTCG	CGGGCTGGG	CGGTATGGCG	GCCGAGGACG
	3901	CCAGCCCGGT	GGGGCGGCCG	CGCCCAAAGTG	CAGTACGCCG	ACCGTGGCCG	GCGGGAGGAC
	3961	CGGACCGGTC	AGTGCAGTCC	CGCGGCCCTG	CGGGACCGCT	CGTCCCAGAC	GGGTTCCACC
	4021	GCGGCGAAC	GGGGTCCGTG	TCCGCGGCCG	TAGACCATCA	GTGTCGCTC	GAAGGTGATG
20	4081	ACGATGACAC	CCTCTGGTT	GTAGCCGATG	GTGCGCACCG	TGATGATGCC	TACGTCAAGT
	4141	CGGCTGGCGG	ACTCCGGGT	GTTCAAGGAC	TCGGACTCG	AGTAGATGGT	GTCGCCCTCG
	4201	AAGACGGGGT	TCGGCAGCCT	GACCCGGTCC	CAGCCGAGGT	TGGCCATCAC	ATGCTGGGAG
	4261	ATGTCGGTGA	CGCTCTGCC	GGTGCAGCAGG	GCGAGGGTGA	AGGTGGAGTC	CACCAGCGC
	4321	TTGCCCCAGG	TGGTCCCCCG	CGAGTAGTGG	CGGTCGAAGT	CGACGGGCCG	GGTGTCTGC
25	4381	GTCAGGAGCG	TGAGCCAGGA	GTTGTCGGTC	TCCAGGACCG	TGCGGCCAG	GGGGTGGCGG
	4441	TACACGTCG	CGGTGGTGA	TCCTCTGAAG	TAGCGGCCCT	GCCAGCCCTC	GACCACAGCG
	4501	GTGCGGGTGG	CGTCCTGGTC	CGGGTTCTCA	GTGTCATGG	CGCTCATTCT	GGGAAGTCCC
	4561	CGGTCGCTG	TGAAATGCCG	AACCTTCACC	GGGCTCATAC	GTGCGCGCA	TGAGCCCTGG
	4621	ACCGTACGTA	GTCTAGAAC	CTCGCCACCA	CTGGCGCGC	TGGTCCCTCCG	GCGAGTGTGA
30	4681	CCACGCCGAC	CGTGCGCCG	GCCTGCGGG	CGTCGAGCGG	CACGGCGACG	GCGTGGTCAC
	4741	GGGGCCCGGA	GGGGCTGCCG	GTGAGGGGGG	CGACGGCCAC	ACCGAGGCCG	GCGCGACCA
	4801	GGGCCCCGAG	CGTGCAGCAG	TCGGTGCCTC	CCAGGACGAC	CGCGGCCACG	AATCCGGCCG
	4861	CGGCGCACAG	CGGTCGGGT	ATCTGGCGA	GTCCGAAGAC	CGGCTCCAGT	GCCACGAACG
	4921	CCTCATCGGC	CAGCTCCGCG	GTCCGCACCC	GGCAGCGTCT	GGCCAGGCCG	TGTCCGGGTG
35	4981	GGACGAGCAG	GCACAGTGCC	TCGTCGGCA	GTGGTGTCCA	CTCCACATCG	TCCCCGGCGG
	5041	GTCGTGGGCT	GGTCAGCCCC	AGGTCCAGCC	TGCTGTTGCG	GACGTCGTG	ACCACGGCGT
	5101	CGGGCGCGTC	GCCGCGCAGT	TCGAAGGGTGG	TGCGGGGAGC	CAGCCGGCGG	TACCCGGCGA
	5161	GGAGGTCGGG	CACCAAGCCAG	GTGCCGTAGG	AGTGCAGGAA	ACCCAGTGC	ACGGTGGCGG
	5221	TGTCGGGGTC	GATCAGGGCG	GTGATGCGCT	GCTCGGCC	GGAGACCTCA	CTGATCGCGC
	5281	GCAGGGCGTG	GGCGCGGAAG	ACCTCGCCGT	ACTTGTGAG	CGGGAGCCGG	TTCTGGTGCC
40	5341	GGTCGAACAG	CGGCACGCC	ACTCGTCGCT	CCAGCCGCCG	GATGGCCCTG	GACAGGGTCG
	5401	GCTGGGAGAT	GTTGAGCCG	TCCGCGGTGA	TCGTCACGTG	CTCGTGTG	GCCAAGGCCG
	5461	TGAACCACTG	CAACTCCCGT	ATCTCCATGC	AGGGACTATA	CGTACCGGGC	ATGGTCTCTGG
	5521	CGAGGTTTCG	TCATTTACA	GCGGCCGGG	GGCGGCCAC	AGTGAAGTCCT	CACCAACCAG
	5581	GACCCCATGG	GAGGGACCCC	ATGTCGAGC	CGCATCCTCG	CCCTGAACAG	GAACGCCCG
45	5641	CGGGCCCCCT	GTCCGGTCTG	CTCGTGGTTT	CTTGAGGCA	GGCCGTGCG	GCTCCGTTG
	5701	CCACCCGCCA	CCTGGCGGAC	CTGGCGCCC	GTGTCATCAA	GATCGAACGC	CCCGCGAGCG
	5761	GCGACCTCGC	CCGCGGCTAC	GACCGCACGG	TGCGTGGCAT	GTCCAGGCCAC	TTCGTCTGGC
	5821	TGAACCGGGG	GAAGGAGAGC	GTCCAGCTCG	ATGTGCGCTC	GCCGGAGGGC	AACCGGCACC
	5881	TGCAAGCCTT	GGTGGACCGG	GCGATGTC	TGGTGCAGAA	TCTGGCACCC	GGCGCCGCGG
50	5941	GCCGCTGGC	ATCGGCCACC	AGGTCCCTCG	GCGGAGCCAC	CGAGGCTGAT	CACCTCGGGA
	6001	CATATCCGGC	TACGGCAGTA	CCGGCTGCTA	CCGCGGACCG	CAAGGCGTAC	GACCTCTGG
	6061	TCCAGTGC	AGCGGGGCTG	GTCTCCATCA	CCGGCACCCC	CGAGAACCCG	TCCAAGGTGG
	6121	GCCTGTCCAT	CGCGGACATC	TGTGCGGGG	TGTACGCGTA	CTCCGGCATC	CTCACGGCCC
	6181	TGCTGAAGCG	GGCCCGCAC	GGCCCCGGCT	CCGAGTTGGA	GGTCTCGATG	CTCGAAGCCC
55	6241	TCGGTGAATG	GATGGGATAC	GCGAGTACT	ACACGCGCTA	CGGGCGCACC	GCTCCGGCCC
	6301	CGCGGGCGC	CAGCCACGCC	ACGATGCC	CCTACGGCCC	GTTACCCACG	CGCGACGGGC
	6361	AGACGATCAA	TCTCGGGCTC	CAGAACGAGC	GGGAGTGGGC	TTCTTCTG	GGTGTGTCG
	6421	TACAACGCC	CGGTCTCTGC	GACGACCCGC	GCTTTCCGG	CAACGCCGAC	CGGGGTGGCG
	6481	ACCGCACCGA	GCTCGACGCC	CTGGTGAGCG	AGGTGACGGG	CACGCTCACC	GGCGAGGAAC
60	6541	TGGTGGCGG	GCTGGAGGAG	GCGTCGATCG	CCTACCGCACG	CCAGCGCACC	GTGGGGAGT
	6601	TCAGCGAAC	CCCCCAACTG	CGTGCACCG	GACGCTGGGC	TCCGTTGAC	AGCCCCTGCG
	6661	GTGCGCTGGA	GGGCTGTAC	CCCCCGGTCA	CCTTCCACGG	CGAGCACCCG	CGGGCGCTGG
	6721	GCCGGGTCCC	GGAGCTGGC	GAGCATACCG	AGTCCGTCCT	GGCGTGGCTG	GCCGCGCCCC
	6781	ACAGGCCGA	CGCGGAAGAG	GCGGCCATG	CCGAATGAAC	TCACCGGAGT	CCTGATCCTG

	6841	GCCGCCGTGT	TCCTGCTCGC	CGGCGTACGG	GGGCTGAACA	TGGGCCTGCT	CGCGCTGGTC
	6901	GCCACCTTC	TGCTCGGGGT	GGTCGCACTC	GACCGAACGC	CGGACGAGGT	GCTGGCGGGT
	6961	TTCCCCCGA	GCATGTTCT	GGTGCCTGGTC	GCCGTCACGT	TCCCTTCGG	GATCGCCCAC
5	7021	GTCAACGGCA	CGGTGGACTG	GCTGGTACGT	GTCGCGGTGC	GGCGGGTGGG	GGCCCGGGTG
	7081	GGAGCCGTCC	CCTGGGTGCT	CTTCGGCCTG	GGCGCACCTG	TCTGCGCGAC	AGGCAGGGCC
	7141	TCGCCCCGCG	CGGTGGCGAT	CGTGGCGCCG	ATCAGCGTCG	CGTCGCCGT	CAGGCACCAC
	7201	ATCGATCCGC	TGTACGCCG	ACTGATGGCG	GTGAACGGGG	CCGAGCCGG	CAGTTTCGCC
	7261	CCCTCCGGGA	TCCTGGGCGG	CATCGTCCAC	TCGGCGCTGG	AGAAGAACCA	TCTGCCCGTC
	7321	AGCGGGGGC	TGCTCTTCGC	AGGCACCTC	GCCTTCAACC	TGGCGGTGCG	CGCGGT-JTCA
10	7381	TGGCTCGTCC	TCGGGCGCAG	CGCCTCGAA	CCACATGACC	TGGACGAGGA	CACCGATCCC
	7441	ACGGAAGGGG	ACCCGGCTC	CCGCCCCGGC	GCGGAACACG	TGATGACGCT	GACCGCGATG
	7501	GCCGCCGTGG	TGCTGGGAAC	CACGGTCCTC	TCCCTGGACA	CCGCTTCCT	GGCCCTCAC
	7561	TTGGCCGGCT	TGCTGGCGCT	GCTCTCCCG	CGCACCTCCC	AGCAGGCCAC	CAAGGAGATC
	7621	GCCTGGCCCG	TGGTGCCTGCT	GGTATGCGGG	ATCGTGACCT	ACGTGCCCT	GCTCCAGGAG
15	7681	CTGGGCATCG	TGGACTCCCT	GGGAAAGATG	ATCGCGGGCA	TCGGCACCCC	GCTGCTGGCC
	7741	GCCCTGGTGA	TCTGCTACGT	GGGCGGTGTC	GTCTCGGCT	TCGCCTCGAC	CACCGGGATC
	7801	CTCGGTGCC	TGATGCCGCT	GTCCGAGCCG	TTCTGAACT	CCGGTGCAT	CGGGACGACC
	7861	GGCATGGTGA	TGGCCCTGGC	GGCCCGGGCG	ACCGTGGTGG	ACGGGAGTCC	CTTCTCCACC
	7921	AATGGTGCTC	TGGTGGTGGC	CAACGCTCCC	GAGCGCTGC	GGCCCGGGCGT	GTACCAGGGG
20	7981	TTGCTGTGGT	GGGGCGCCGG	GGTGTGCGCA	CTGGCTCCC	CGGGCGCCTG	GGCGGCCCTC
	8041	GTGGTGGCGT	GAGCGCAGCG	GAGCGGGAAAT	CCCCTGGAGC	CGGTTCCCG	TGCTGTGTCG
	8101	CTGACGTAGC	GTCAAGTCCA	CGTGCCGGGC	GGGCAGTAGC	CCTAGCATGT	CGGGCATGGC
	8161	TAATCAGATA	ACCCGTCCG	ACACGCTGCT	CGTTACGTA	CGGAAGGTGT	CCCTGCGCGA
	8221	TGACGAGGTG	CTGAGGCCG	TGCGCGCGA	GACGGCCGAG	CTGCCGGCG	GTGGCGTACT
25	8281	GCCGGTGCAG	GGCGAGGAGG	GACAGTTCC	CGAGTTCTG	GTGCGGTTGA	CCGGCGCGC
	8341	TCAGGTGCTG	GAGATCGGG	CGTACACC	CTACAGCACG	CTCTGCC	CCCGCGGATT
	8401	GGCGCCCGGG	GGCCGTGTTG	TGACGTGCGA	TGTCATGCC	AAGTGGCCCG	AGGTGGCGA
	8461	GCGGTACTGG	GAGGAGGCCG	GGGTTGCCGA	CCGGATGCC	GTCCGGATCG	GCGACGCCG
	8521	GACCGTCCTC	ACCGGGCTGC	TCGACGAGGC	GGGCGCGGGG	CCGGAGTCGT	TCGACATGGT
30	8581	GTTCATCGAC	GGCGACAAGG	CCGGTACCC	CGCCTACTAC	GAGGCGGCC	TGCGCTGGT
	8641	ACGCCGCGGC	GGGCTGATCG	TCGTCGACAA	CACGCTGTC	TTCGGCCGGG	TGGCCGACGA
	8701	AGCGGTGCAG	GACCGGACA	CGGTCGCGG	ACCGAAGTC	AACCGGGCAC	TGCGCGACGA
	8761	CGACGGGGT	GACCTGGCGA	TGCTGACGAC	GGCCGACGGC	GTCACCC	TGCGGAAACG
	8821	GTGACGGGGG	CGATGTCGGC	GGCGTGCAGC	GTCAGCGTC	TCGGCGCGGG	CCTCGCGGAG
35	8881	GGCTCCAGAT	GCAGGCCTGC	GACGCCGGCG	GCGGAAGCGC	CCGCCACCTC	GGACACCGAG
	8941	GGGCAGTCGG	AGTCCCGCAA	GCCCGCGAAC	CGGTAGGCGA	TCTCCATCAT	CGGGTTGCGG
	9001	TCCGTACGCC	GGAAGTCCG	CACCAAGTGC	GCCCCCGCGC	GGGCGCCCTG	GTCCGTGAGC
	9061	CAGTTCAAGGA	TCGTCGACC	GGCACCGAAC	GACACGACCC	GGCAGGACGT	GGCGAGCAGT
	9121	TTCAGGTGCC	ACGTCGACGG	CTTCTCTCC	AGCAGGATGA	TGCCGACGGC	GGCGTGCAGG
40	9181	CCGAAGCGGT	CGCCCCATGGT	GACGACGAGG	ACCTCATGGG	CGGGATCGGT	GAGCACCGC
	9241	GCAGGTGCGC	GTCGGAGTAG	TGACGCC	TCGCGTTCAT	CTGGCTGGTC	CGCAGCGTC
	9301	GTTCTCGAC	GGGGCTGAGT	TCCCTCTCCC	CCGCGGGTGC	GATCGTCATG	GAGAGGTGCA
	9361	GCGAGCGCAG	GAAGTCCCTG	TCGGGACCGG	AGTACGCC	CCGGGCTGG	TCGCGCGCAG
	9421	AACCCGCTG	GTACATCAGG	CGGCCGCGAC	GCGAGTCAC	CGTGGACACC	GGCGGGCTGA
45	9481	ACTCCGGCAG	CGACAGGAGC	GTGGCCGCCT	GCTCGGCC	GTAGCACCGC	ACCTCGGGCA
	9541	GGTGGAACGC	CACCTCGGCC	CGCTCGGCC	GCTGGTCGTC	GATGAACCGC	ATCGTGGTCG
	9601	GTGCGGAAGTT	CAGCTCCGTG	GCGATCTGC	GGACGGACGT	CGACTTCGGC	CCCCATCCGA
	9661	TGCGGGCCAG	CACGAAGTAC	TCCGCCACAC	CGAGGCCTTC	CAGACGCTCC	CACGCGAGGT
	9721	CGTGGTCGTT	CTTGTCTGCC	ACCGCC	GGATGCC	GTGCGCAGC	GTGGTGATCA
50	9781	CCTCGCGGAT	CTCGTCGGT	AGGACCAC	CGTCGCTC	CAGCACGGTG	CCCCGCCACA
	9841	AGGTGGTGTG	CAAGTCCCAG	ACCAAGAC	TGACAA	CATGGCTGTC	CTCTCAAGCC
	9901	GGGAGCGCA	GGCGCTGCTG	GGCCGACATC	ACCCGGCAC	TCTCGCTGCT	GGCCCTCGATG
	9961	ATCTCCATGA	GCTTGGCGT	GGCGTACGCC	CGTCGACGA	CGTGTCCCTC	TCTCGCGCCT
	10021	GCCGACCGCA	GCACCTGTGC	GGCGTGC	GCCCCGGCG	GGGCTCGTTC	GGCGCGCAGG
55	10081	TGCTGGCCA	GGATCGTCG	GGGACCCATC	TGGGCGAGC	CCTCGTCCC	GTGGTCGCTG
	10141	GCGTACTCGC	ACACGCGGGC	CGCGATCTG	TCCGCGGTCC	ACAGGTGCGC	GATGTGCCCG
	10201	GCGACGAGTT	GGTGGTCGCC	GAGCGGCC	CCGAAGTCGT	CCCGGGTCCG	GGCGTGGGCC
	10261	ACCGCGGCCG	TGCGGCAGGC	CCGCAGGATC	CCGACGCGAC	CCCAGGCCAC	CGACTTGCAC
	10321	CCGTAGGCCA	GTGACGCC	GACCA	GGCAGTGA	CGCCGGAGCC	GGCCAGGACC
60	10381	GCGCCGGCCG	GCACACGCA	CTGGTCCAGG	TGCA	CGTGGCCGGC	GGCGCGCAG
	10441	CCGGACGGCT	TCGGGACGCG	CTCGACGCGT	ACGCCGGGG	TGTCGGCCGGG	CACGACCA
	10501	ACCGCACCGG	AACCATCC	CTGGAGACCG	AAGACGACCA	GGTGGTCCGC	GTAGGCGCG
	10561	GCAGTCGTCC	AGACCTTGTG	GGCGTGCAGC	ACAGCGGTGT	CCCCGTCGAG	CCGAACCCGC
	10621	GTCCGCATCG	CCGACAGATC	GCTGCCG	TGCGCTCAC	TGAAGCCGAC	GGCGCGAGT

10681 T'TCCCGCTGG TCAGCTCC TT CAGGAAGGTC GCCC GCTGAC CGGC GTCGCC GAGCC GCTGC
 10741 ACGGTCCACG CGGCCATGCC CTGCGACGTC ATGACACTGC GCAGCGA ACT GCAGAGGCTG
 10801 CCGACGTGTG CGGTGAACCTC GCCGTTCTCC CGGCTGCCGA GTCCCAGACC GCCGTGCTCG
 10861 GCGCCCACTT CCGCGCAGAG CAGGCCGTCG GCGCCGAGCC GGACGAGCAG GTCGCGCGC
 5 10921 AGTTGCCGG ACGTGTCCTCA CTCGGCGGCC CGGTACCGA CAAGGTCGGT CAGCAGCGC
 10981 TCACGCTCAG GCATCGACGG CCCCGAGCCG GTGGACGAGT GCGACCATGG ACTCGACGGT
 11041 ACGGAAGTTC GCGAGCTGGA GGTCCGGGCC GGCGATCGT ACGTCGAACG TCTTCTCCAG
 11101 GTACACGACC AGTTCCATCG CGAACAGCGA CGTGAGGCCG CCTTCCCGA ACAGGTCGCG
 11161 GTCCACGGGC CAGTCCGACC TGGTCTTCGT CTTGAGGAAC GCGACCAACG CGTGCACGAC
 10 11221 GGGGTCGTCC TTGACGGGTG CGGTATGAG AACACCTCT CGTATTGTA GAAGCCCGG
 11281 CCGGTCTTCC GGCGTGGTG TCCCTCGCG ACCTTGCCA GCAGCAGGTC ACAGGGCGG
 11341 CTGCGCTCGT CGCCGGTGC G TTTGTGCAGC ACCCACAGCG CGTGCACGAG GTTGTGATG
 11401 CCGATCAGGT CGCGGTGC CAGCGGCCCG GTCGGATGGC CGAGGCACCC CGTCATGAGC
 11461 GCGTCGACGT CCTCGACCGA CGCGGTGCC CGCTGCACGA TCCCGGCCGC GTCGTTGATC
 15 11521 ATCGGGTGG A GCAGCCGGCT CGTGACGAAG CGGGGCGCGT CCCGGACGAC GATCGGCTTG
 11581 CGCCGCAGCG CCGCGAGCAG GTCCCCGGCG GCGGCCATGG CCTTCTCACC GGTCCGGGGT
 11641 CCGCGGATCA CCTCGACCGT CGGGATCAGG TAGCAGGGT TCATGAAGTG CGTGCACGAGC
 11701 AGGTCCCTCGG GCCGGGCCAC GGAGTCGGCC AGTTCGTCAA CCGGGATCGA CGACGTGTC
 11761 GTGATGACCG GGATACCGG CGCCGCTGCC GAGACCGTGG CGAGTACCTC CGCCTTGACC
 20 11821 TCGGCGTCCT CGACGACGGC CTGGATCACC GCGGTGGCCG TACCGATCGC GGGCAGCGC
 11881 GACGTGGCCG TCCGCAGCAC ACCGGGGTCG GCCTCGGCCG GCGCGGCCAC GAGTTGTGCC
 11941 GTCCGCAGTT CGGTGGCGAT CGCGGCCCGC GCGCCGCTAA GGATCTCTC GGACGTGTC
 12001 ACGAGTGTCA CGGGGACGCC GTGGCGCAGC GCGAGCGTGG TGATGCCGGT GCCCACACT
 12061 CCCGCGCCGA GCACGATCAG CTGGTGGTCC ACGCTGTTT CTCCTCCCG GGTCAACATG
 25 12121 GCAGCGAGTA CGGGTGCAGG ACAGTCTTCG GGGTCGACCC GATCGCGTCC TTGCGGCCGA
 12181 GGCGGAGTT GTCGGCGAAG CCGAGCAGCA CGTCAACCG GATGTGGTCG GCGAACGCG
 12241 TGCCCCGTCG A GTCGAGGACG CTCAAGCTGT CCGGGTGGTC CGCCGGCCGGT TCCGGTGGCC
 12301 CGCACAGGGC CGCCAGGAC GGGCGAGCT CGGGTCCGG CAGTTGCTGG TACTCGCCCT
 12361 CGGCGCGGGC CTGCCCCGG A TGTCGACGC AGATGAACGC GTCGTCGAGC AGGGTCTTCG
 30 12421 GCAGTTCGGT CTTGCCCCGGC TCCTGGCGC CGATGGCGTT CACATGCAGG TGCGGCAGCC
 12481 GCGGCTCGGC CGGCAGCACC GGGCTTTGCG CCGAGGGCAC CGAGGTGACG GTGGACAGGA
 12541 CATCCCGGGC GGCGGCGGCC TCGCCGGAT CGGTACCTT GACCGGCAGT CCGAGGAACG
 12601 CGATGCGGTC CGCGAACCGAC GCGCGTGGC CGGGGTCGGT GTCGCTGACC AGGATCCGCT
 12661 CGATGGCAG GACCCCTGCT AGCGCGTGC CGCTGGTCAC CGCCTGTGCC CCCGCGCCGA
 35 12721 TCAGCGTGAG CGTGGCGCTG TCGGACCGGG CCAGCAGCCG GCTCGCAGC GCGGCCACCG
 12781 CGCCGGTCCG CATCGCGGT ATCACGCC TG CGTGGCGAG GGGTCACCTT GACCGGCAGT CCGAGGAACG
 12841 CGTCGTCGAG CGCGCAGCATC GTGCCGACGA CGGGGTCGGT GTCGCTGACC AGGATCCGCT
 12901 GCGGACTGTA CGAAACCGTC TTCATGGTCA CGCCGACACC GGGGACCCGG TACGGCATGA
 12961 ACTCGATGAC GCCGGGAATG TCGCCGCCGC GGACGAATCC GGTACGCGGC GGCGCCCTCG
 40 13021 CGAACTCGCC GCGGCCGAGC GCGCGAACCC CGTCGTCAG CTCGCTGATC AGCCGGTCCA
 13081 TCATCACGTC CGGGCGATC ACGGAGAGAA TCCGCTTGAT GTACGTTGG CGCAGGACCC
 13141 TGGTCTGCAT GTGTCACTTC CCTTCTGTG CGGGAGCTGT CTTGGTGGTG CGCTCGGGG
 13201 CGGCTCCGT TCTCATCGCA GCTCCCTGTC GATGAGGTG AAAATCTCGT CGCGGGTCGC
 13261 GTCCCGGGAC AGCACGCCGG CGGGCGTGGT CGGGCGGGTC TCCCGCCGCC AGCGGTTGAG
 45 13321 CAGGCCGTCC AGCCGGGTT CGATCGCGTC CGCCTGGCG GCGCTGGCG GCGCCCGGGT CGACACCGGC
 13381 AACGAGTGCT TCCAGCCGGT CGAGCTGCGC GAGCACACG GTACCGGGT CGTCCGGGGA
 13441 CAGCAGTTCA CGGATGCGGT CGGCAGTGC GCGCGGCAC GGGTAGTCGA AGACGAGCGT
 13501 GGC GGACAGT CGCAGACCGG TCGCCTCGTT GAGGCCGTT CGCAGCTGCA CGCGATGAG
 13561 CGAGTCCACA CCGAGTTCCC GGAACGCCGC GTCCCTCCGG ATGTCTCCCG GGTGGCGTGC
 50 13621 GCGCCAGACG GCCGCTGCC TCTGCCGGAC GAGGGCGAGC AGGTGGTGG GGC GTTCCCTG
 13681 CTCGTTGCCG GCGCTCCGGC GGGCGACGG CTTGGGCCGG CCACGCAGCA CGGGGAGGTC
 13741 CGGCGGCAGG TCGCCCCGCC CGCGACGAC ACTGCCGTT CCGGTGTGGA CGCGGCCGTC
 13801 GTACATGCCG ATGCCCTGTT CGGCGGTGAG CGCGCTCGCC CCACCCCTTG GCATACGGCG
 13861 CGGGTCGGCG TCGGTCAAGGT CGCGGTCA CGCACTCGCC TGGTCCCACA GCCCCCACGC
 55 13921 GATCGACAGC CCTGGCAGCC CTTGTGCACG CGGGTGTTCG CGGAGCGCGT CGAGGAACGC
 13981 GTTCGCCGCC GCGTAGTTG CCTGACCGGG GGTGCCAGC ACACCGGCCG CCGACGAGTA
 14041 GACGACGAAT CGGGCGAGGT CGGTGTGCG GGTGAGGCCG TGCAAGGTGCC AGGCAGCGTC
 14101 GGCCTTGGGT TTGAGGACGG TGTCGATGCG GTCGGGGGT AGGTTGTGCA GCAGGGCGTC
 14161 GTCGAGGGTT CGGGCGGTGT GGAAGACGGC GGTGAGGGT TGAGGGATGT GGGCGAGGGT
 60 14221 GGTGGCGAGT TGGTGGGGT CGCCGACGTC GCAGGGGAGG TGGGTGCCGG GGGTGGTGT
 14281 GGGGGGTGGG GTGCGGGGAGA GGAGGTAGGT GTGGGGGTGG TTCAGGTGGC GGGCGAGGAT
 14341 GCGCGCGAGG GTGCCGGAGC CGCCGGTGT GACGACGGCC CCTCGGGGGT CCAGGGCCG
 14401 CGGGACCGTG AGGACGATCT TGCCGGTGT CTGCGCCCGG CTCAATGGTC CCAGGCCCTC
 14461 CGGGACCTGC CGCATGTCGT GCACCGTCAC CGGCAGCGGG TGCAGCACAC CGCGGCCGAA

	14521	CAGGCCGAGC	AGCTCCGCGA	TGATCTCCTT	GAGCCGGTCG	GGCCCCGCGT	CCATCAGGTC
	14581	GAACGGTCGC	TGGACGGCGT	GCCGGATGTC	CGTCTTCCCC	ATCTGGATGA	ACCGGCCACC
	14641	CGGCGCGAGC	AGGCCGACGG	ACGGCTCGAG	GAGTTCACCG	GTGAGCGAGT	TGAGCACGAC
5	14701	GTCGACCGGC	GGGAACCGGT	CGGCGAACGC	GGTGCTGCGG	GAATCGGCCA	GATGCGCTCC
	14761	GTCCAGGTCC	ACCAGATGGC	GCTTCGCGGC	GCTGGTGGTC	GCGTACACCT	CCGCGCCCAG
	14821	GTGCCGCGCG	ATCTGCCGGG	CGGCCGAACC	GACACCGCCG	GTGGCCGCGT	GGATCAGGAC
	14881	CTTCTCGCCG	GGGCGCAGCC	CGGCGAGGTC	GACCAGGCCG	TACCAACGCG	TCGCGAACGC
	14941	GGTCATCACG	GACGCCGCCT	GCGGGAACGT	CCAGCCGTCC	GGCATCCGGC	CGAGCATCCG
10	15001	GTGGTCGGCG	ATGACCGTGG	GGCCGAAGCC	GGTGCCGACG	AGGCCGAAGA	CGCGGTCGCC
	15061	CGGTGCCAGA	CGGGAGACGT	CGGCCCGGT	CTCCAGGACG	ATGCCCGCGG	CCTCGCCGCC
	15121	GAGCACGCC	TGACCGGGGT	AGGTGCCGAG	CGCGATCAGC	ACATCGCGGA	AGTTGAGGCC
	15181	CGCCGCACGC	ACACCGATCC	GGACCTCGGC	CGGGCGAGG	GGCGCCGGG	GCTCCGCCGA
	15241	GTCGGCGCG	GTGAGGCGT	CGAGGGTGC	CGTCCGCGCC	GGCCGGATCA	GCCACGTGTC
15	15301	GCTGTCGCCG	ACGGTGAGCG	GCTCCGGCAC	CCGGGTGAGG	CGGCCGCGCT	CGAACCGGCC
	15361	GCCCGCGCAGC	CGCAGACCGC	GCTCGCCGAG	TGCGACGGCG	ATGCGCTGCT	GCTCGGGGGC
	15421	GAGCGTGACG	CGGGAACCTCGG	TCTGACGCTG	GACGAACCGG	CGGGCTGCTG	CGGCCTGGGC
	15481	GGCGCGCAGC	AGTCCGGCGC	CCGCGCCGGT	GGCGAGGCC	CGGGTGGTGT	GCACGAGCAG
	15541	ATCCC CGCCG	GAGCCGGTCA	GGGGTGTAG	CAGCCGGGTG	GTGAGCGCAC	GCGTCTCGGC
20	15601	CACCGGGTCG	TCGCCATCAG	CGGCAGGCAA	CGTGTATGACG	TCCACGTCGG	TCGCGGGGAC
	15661	ATCCGTGGGT	GGGGCGACCT	CGATCCAGGT	GAGACGCATC	AGGCCGGTGC	CGACGGGTGG
	15721	GGACAGCGGG	CGGGTGC	CCGTCGGAT	CTCGCGACG	AGTTGGCCGG	CGGAGTCGGC
	15781	GACGCCAGA	CTCAGCTCGT	CGCCCGTAC	AGTGTATCAG	GCTCGGAGCA	TGGCGAGCC
	15841	CGTGGCGAGC	AACCGGGCCC	CCTTCAGGC	GAACGGCAGA	CCCGCAGCGC	TGTCGTCGG
25	15901	CGTGGTGAGG	GGCACGGCGT	GCAGGGCCG	GTCGAGCAGC	GCCGGATGCA	CACCGAAACC
	15961	GTCCGCCCTCG	GGGGCCTGCT	CGTCCGGCAG	CGCCACCTCG	GCATAACACGG	TGTCACCATC
	16021	ACGCCAGGCA	GGCCGCAACC	CCTGGAACGC	CGACCGTAC	TCTAACCGG	CATCCCCAG
	16081	TTCGTATAG	AACCCCGAGA	CGTGCAGCGC	CACGGCCGTG	ACCGGCGGCC	ACTGCGAGAA
	16141	CGGCTCCACA	CCGACAAACAC	CGGGGGTGTG	GGGGGTGTG	GGGGTCAGGG	TGCCGCTGGC
	16201	GTGCCGGGTC	CAGCTGCCG	TGCCCTCGGT	ACGCGCTGG	ACGGTCACCG	GCCGCCGTCC
30	16261	GGCCTCATCA	GGCCCTTCCA	CGGTACCGA	CACATCCACC	GCTCGGTCA	CCGGCACCA
	16321	AAAGGGGGAT	TCGATGACCA	GCTCGTCCAC	TATCCCGAA	CCGGTCTCGT	CACCGGGCCG
	16381	GATGACCAGC	TCCACAAACG	CCGTACCCGG	CAGCAGGACC	GTGCCCCGCA	CCGCGTGTAC
	16441	AGCCAGCCAG	GGGTGAGTGC	GCAATGAGAT	CCGGCCAGTG	AGAACAAACAC	CACCATCGTC
	16501	GGCGGGCAGC	GCTGTGACAG	CGGCCAGCAT	CGGATGCC	GCACCCGTCA	ACCCCGCCGC
35	16561	CGACAGATCG	GTGGCACCGG	CCGCCCTCCAG	CCAGTACCGC	CTGTGCTCGA	ACCGGTACGT
	16621	GGGCAGATCC	AGCAGCCGTC	CCGGCACCGG	TTCGACCACC	GTGCCCCAGT	CCACTGCCGT
	16681	GCCCAGGGTC	CACGCCCTCG	CCAACGCCGT	CAGCCACCGC	TCCCAGGCC	CGTCACCGGT
	16741	CCGCAACGAC	GCCACCGTGT	GAGCCTGCTC	CATCGCCGGC	AGCAGCACCG	GATGGGACT
	16801	GCACTCCACG	AACACCGACC	CATCCAGCTC	CGCCACCGCC	GCGTCCAACG	CCACCGGACG
40	16861	ACGCAGATT	CGGTACCGA	ACCCCTCATC	CACCGGCTCC	GTCACCCAGG	CGCTGTCAC
	16921	GGTCGACCAC	CACGCCACCG	ACGCGGCC	CCCTGCCACC	CCCTCCAGTA	CCTTGGCCAG
	16981	TTCATCCTCG	ATGGCTTCA	CGTGGGGCGT	GTGGGAGGCG	TAGTCGACCG	CGATACGACG
	17041	CACCGCAGC	CCTCGGCC	CATACCGC	CACCCACCTCC	TCCACCGCC	ACGGGTCCCC
	17101	CGCCACCA	GTCGAAGCGC	GGCCGTTACG	CGCCGCGATC	CACACACCC	CGACCAAGAC
45	17161	GACCTCACCG	GGCGCAACG	CCACCGAAGC	CATCGCTCC	CGCCCGGCC	GTCGCGCCGC
	17221	GATGACCTGA	CTGCGCAATG	CCACCAACGCG	GGCGCGTCC	TCGAGGCTGA	GGGCTCCGGC
	17281	CACGCACGCC	GGCGCGATCT	CGCCCTGGGA	GTGTCGATC	ACCGCGTCCG	GCACGACCCC
	17341	ATGCGCCTGC	CACAGCGCGG	CCAGGCTCAC	CGCGACCGCC	CAGCTGGCCG	GCTGGACCAC
	17401	CTCCACCCGC	TCCGCCACAT	CCGGCCGCGC	CAACATCTC	CGCACATCCC	AGCCCCGTG
50	17461	CGGCAGAAC	GCCTGAGCGC	ACTCCTCCAT	ACGCGCGC	AACACCGCGG	AGTGGGACAT
	17521	GAGTTCCACG	CCCATGCCG	CCCATGGGG	GCCCTGGGG	GGGAAGACGA	ACACCGTACG
	17581	CGECTGGTCC	ACCGCCACAC	CCGTCACCCG	GGCATCGCC	AGCAGCACCG	CACGGTACCC
	17641	GAAGACAGCA	CGCTCCCGCA	CCAACCCCTG	CGCGACCGCG	GCCACATCCA	CACCAACCCC
	17701	GCGCAGATA	CCCTCCAGCC	GCTCCACCTG	CCCCCGCAGA	CTCACCTCAC	CACGAGCCGA
55	17761	CACCGGCAAC	GGCACCAAC	CGTCAACAAAC	CGACTCCCCA	CGCGACGGCC	CAGGAACACC
	17821	CTCAAGGATC	ACGTGCGCGT	TCGTACCGCT	CACCCCGAAC	GACGACACAC	CCGCATGCGG
	17881	TGCCCCATCC	GAATCGGGCC	ACGGCC	CTCGGTGAGC	AGCTCCACCG	CACCGGGCGA
	17941	CCAGTCCACA	TGCGACGACG	GCTCGTCCAC	ATGAGCGTC	TTCCGGCGCA	TCCCCTACCG
	18001	CATGCCCATG	ACCATCTTGA	TCACACCGGC	GACACCCGCC	GCCGCC	CATGACCGAT
60	18061	GTTGACTTC	AAAGAACCCA	GCAGCAGCGG	AACCTCACGC	TCCTGCCCGT	ACGTGCCCAG
	18121	AATGGCCCTGC	GCCTCGATGG	GATCGCCCGAG	CGTCGCCCC	GTCCCCTGCG	CCTCCACCC
	18181	GTCCACATCG	GGCGCGCGCA	GTCCGGCGTT	CACCAACGCC	TGCTGGATGA	CACGCTGCTG
	18241	GGACGGGCCG	TTGGGGCGG	ACAGCCGTT	GGAGGACCG	TCCCTGGTCA	CCGCCGACCC
	18301	GCAGGACGACC	GGCAGAACCGG	TGTGCGTGT	GCGCTCGCG	TGGAGAGGCC	GCTCCAGCAC

18361 AAGAACGCCG GCGCCCTCCG CCCAGCCGGT GCCGTTGGCG GCGTCGGCGA ACGCGCGGCA
 18421 GCGGCCGTG GGGGAGAGTC CGCCCTGCTG CTGGAAATTCC ACGAACCCGG TCGGGGTGCG
 18481 CATGACGGTG ACACCGCCGA CCAGCGCCAG CGAGCACCTC CCGTGGCGCA GTGCGTCCC
 18541 GGCCTGGTGC AGCGCGACCA GCGACGACGA GCACGCCGTG TCCACCGTGA ACGCCGGTCC
 5 18601 CTGGAGCCCC TAGAAGTACG AGATCCGGCC GGTGAGCACG CTGGGCTGCA TGCCGATCGA
 18661 GCGGAACCCG TCCAGGTCCG CGCCGACGCC GTACCCGTAC GAGAAGGCGC CCATGAACAC
 18721 GCGGGTGTG CTGCCGCGA GTGTGCCCAG CACGATGCC GCGCTCTCGA ACGCTCCCA
 18781 TGTCGTTTCC AGCAGGATCC GCTGCTGGGG GTCCATGCC CGTGCTCTCAC GGGGGCTGAT
 18841 GCGGAAGAAC GCGGCATCGA AGCCGGCGGC GTCGGAGAGG AAGCCGCCGC GGTCCGTGTC
 10 18901 CGATCCGCCG GTGAGGCCGG ACGGGTCCCCA GCCACGGTCG GCGGGGAAGC CGGTGACCGC
 18961 GTCGCCGCCA CTGTCCACCA TGCGCCACAG GTCTGCCGGC GAGGTGACGC CGCCCGGCAG
 19021 TCGGCAGGCC ATGCCACCGA TGGCCAGCGG TTCTGCCACGG GTCCGGCGG CTGTGGGAAC
 19081 AGCGACCGGT GCGGCACCCAC CGACCAGAGC CTCTGCCAAC CGCGACGCCA TGGCCCGCGG
 19141 CGTCGGGTAG TCGAAGACAA GCGTGGCGGG CAGTCGGACA CGGTCGCCG CGGCGAGTCG
 15 19201 GTTCCGCAGT TCGACGGCGG TCAGCGAGTC GATAACCAAGT TCCCTGAAGG CGCGCTCCGC
 19261 GGACACGTCC GCGGCCGTCC CGTGGCCAG CACCGCCGC GCGTTGTCGC GGACCAAGTGC
 19321 CAGCAGCGCG GTGTCCCCTG CAGGCCCGGA CATGGTGGCG AGCCGGTCGG CGAGCGAAC
 19381 GCGGGTGGCC GCGCCGGGGC GCGATACGCC CGGGCGCAGA TCGGCAGAAA CGGGCGATGT
 20 19441 GTGCGGGGTG AGGTCCATCG TGGCCGCCAC GCGAAGCGC GTGCCGGTTC CGGCCGCCGC
 19501 TTCCAGCAGG CGCATGCCCA CACCGGCCGA CATGGGGCGG AAACCGCCGC GGCGGACACG
 19561 GGTGCGGTG GTGCCGCTCA TGCTGCCGGT GAGTCCGCTG TCATCGGCC AGAGGCCCCA
 19621 GGCCAGCGAC AGCGCGGGCA GTCCCTCGGC ATGGCGCAGC GTCCGAGTC CGTCAAGGAA
 19681 CCCGTTCGCC GCGAGTAGT TGCCCTGGCC GCGGCCGCCG ATGATGCCCG CGACGGACGA
 19741 GTAGAGGACG AACGAGCGCA GGTCCCGCAGC CGGGGTCAAGC TCGTGCAGGT GCCAGGCAGC
 25 19801 GTCGGTTTG GGGCGCAGTG TGGTGGCGAG CGCCTCCGGG GTGAGTGCCG TGGTCACGCC
 19861 GTCGTCGAGC ACGGCTGCCG TGTGAAAGAC CGCGTGTGAGC GGCTGCCGG CGGCCGGCAG
 19921 CGCGCGGCCG AGCTGGTCCC GGTGGCGAC GTCACAGCGG ATGTTGGACAC CGGGAGTGT
 19981 CGCCGGCGGT TCGCTGCCG ACAGCAACAG GAGGTGGCGG GCGGCATGCT CGGCGACGAG
 20041 ATGCCGGCG AGGAGACCTG CCAGCACACC CGAGCCGCCG GTGATGCCCA CGGTGCCGTC
 30 20101 CGGGTCGAGC AGCGGTTCCG GCGTTTCCGC GCGGCCCGTG CGGGTGAACC CGGGCGCTTC
 20161 GTACCGGCCG TCGGTGACGC GGACGTACGG CTGGCCAGT GTCTGGCGG CGGCCAGCCC
 20221 CTCGATGGGG GTGTGGTGC CGGTCTCCAC CAGCACGAAC CGGGCCGGGT GCTCGGCTG
 20281 GGCAGGACCGG ACGAGGCCGG CGACCGCTCC TCCGACCGGT CCCGCTGCGA TCCGGACGAC
 20341 GACGGTGGTC TCCGCAAGGGC CGTCCCTCGGC GATCACCCGG TGCAGCTCGC CGAGCACGAA
 35 20401 CTCGGTGAGC CGGTACGTTCT CGTCGAGGAC ATCCGCCCGG GGTCCGGGA CGCGGGAGAC
 20461 GATGTGGACC GCGTCCGCA GACCGGGCCC GGGAGTGGC AGCTCGGTC AGGAGAGGCC
 20521 GTACAAGGAG TTCCGTAAGA CGGCGGCCAGC GCGTCGACG TTCACCGGTC GCGCGGTCA
 20581 CGCGGGCGACG GTCAACCACCG GTTGGCCGAC CGGGTCCGTC GCATGCACGG CAGGCCGTC
 20641 CGGGCCCTGA GTGATCGTGA CGGCCAGCGT GGTGGCCCG GTCTGTGGA ACCGCACGCC
 40 20701 GCTCCACGAG AACGGCAGCC GCACCTCCGC TTCTGTTC GCGAGCAAGCG GCAGGCAGGT
 20761 GACGTGCAAG GCGCGTCTGA ACAGCGCCGG GTGGACGCCA TAGTGCAGCG TGTCTGCCG
 20821 CTGTTCCCCG GCGATCTCCA CCTCGCGTA CAGGGTTTCG CGTCGCGGCC AGGCGGTGCG
 20881 CAGTCCTGG AACGCTGGGC CGTAGCTGTA GCCGGTCTCG GCGAGCCGCT CGTAGAACGC
 20941 GCTCACGTCG ACGCGTCGCC CGCCCGCGG CGGCCACGCG GCGGGCGGGG CGGCCGCGAC
 45 21001 GCTTCGGCC CGGCCGAGGG TGCCGCTGGC GTGCCGGTC CAGCTGTCCG TGCCCTCGGT
 21061 ACGCGCGTGG ACGGTCACTC GCCGCCGTCC GGCCCTATCG GCCCCTTCGA CGGTACCGA
 21121 CACATCCACC GCGCCGGTCA CGGCCACAC GAGCGGGGTC TCGATGCCCA GTTCATCCAC
 21181 CACCCCGCAA CGGGTCTCGT CACCGGCCGG GATGACCGAG TCCACAAACG CCGTACCGG
 21241 CAGCAGAACCG GTGCCCGCAGA CGCGTGTATC AGCCAGCCAG GGATGCGTAC GCAACGAGAT
 50 21301 CGGGCAGTG AGAACAAACAC CACCAACCGC GTCCGGGGC AGTGTGTGA CGGCCGCCAG
 21361 CATCGGATGC GCGCCCGGG TCAGGCCGGC CGCGGACAGA TCGGTGGCAC CGGCCGCTC
 21421 CAGCCAGTAC CGCCTGTGCT CGAACCGCTA GGTGGGAGA TCGAGCAGCC GTCCCGGCAC
 21481 CGGTCGACC ACCGTGTCCC AGTCCACTGC CGTGGCCAGG GTCCACGCC GCGCCAACGC
 21541 CGTCAGCCAC CGCTCCCGAC CGCCGTCACC GGTCCGCAAC GACGCCACCG TGTGAGCTG
 55 21601 TTCCATCGCC GGCAGCAGCA CGGGATGGGC GTCGACTCC ACGAACACCG ACCCGTCCAG
 21661 CTCCGCCACC GCGCGTCTA GCGCGACGGG GCGACGCAAGG TTCCGGTACC AGTAGCCCTC
 21721 ATCCACCGGC TCGGTACCC AGGGCGTGTG CACCGTGGAC CACCAAGGCCA CCGACCCGGT
 21781 CCCGCCGAA ATCCCCCTCA GTACCTCGGC CAACTCGTC TCGATGGCTT CCACGTGGGG
 21841 CGTGTGGAG CGTAGCTGA CGCGATACAG GCGCACTCGC ACGCTTCGG CCTCGTACCG
 60 21901 CGTCACCACT TCTTCCACCG CGGACGGGTC CCCCCGCCAC ACAGTCGAAG ACGGGCCGTT
 21961 ACGCGCCGCG ATCCACACGC CCTCGACAG GTCCACCTCA CGGCCCGGCA ACGCCACCGA
 22021 AGCCATCGCC CCCGCCCGG CGAGCCGCC GCGGATCACCG TGGCTGCGCA AGGCCACAC
 22081 CGGGCGGGCG TCCTCAAGGC TGAGGGCTCC GGCCACACAC GCGCCGCCGA TCTCGCCCTG
 22141 GGAGTGTCCG ACCACCGCGT CGGGCACGAC CCCATGCCG TGCCACAGCG CGGCCAGGGCT

	22201	CACCGCGGACC	GCCCAGCTGG	CCGGCTGGAC	CACCTCCACC	CGCTCCGCCA	CATCCGGCCG
	22261	CGCCAACATC	TCCCCGCACAT	CCCAGCCCGT	GTGCGGCAAC	AACGCCCGCG	CACACTCCTC
	22321	CATAAGGAGCC	GCAGAACACCG	CAGAACACGC	CATCAACTCC	ACACCCATGC	CCACCCACTG
5	22381	AGCACCCCTGC	CCGGGAAAGA	CGAACACCGT	ACGCGGCTGA	TCCACCGCCA	CACCCATCAC
	22441	CGGGGCATCG	CCCAACAACA	CCGCACGGTG	ACCGAAGACA	GCACGCTCAC	GCACCAACCC
	22501	CTGCGCGACC	GGGGCCACAT	CCACACCACC	CCCGCGCAGA	TACCCCTCCA	GCCGCTCCAC
	22561	CTGCCCCCGC	AGACTCACCT	CACTCCGAGC	CGACACCGGC	AACGGCACCA	ACCCATCGAC
	22621	AGCCGACTCC	CCACCGGACG	GCCCCGGAAC	ACCCCTCAAGG	ATCACGTGCG	CGTTCGTACC
	22681	GCTCACCCCG	AAAGCGGAGA	CACCGGCCCC	GCGCGGACGT	CCCGCGTCGG	GCCACGCCCG
10	22741	CGCCTCGGTG	AGCAGTTCCA	CCGCGCCCTC	GGTCCAGTCC	ACATGCGACG	ACGGCTCGTC
	22801	CACATGCAGC	GTCTTCGGCG	CGATGCCATA	CCGCATCGCC	ATGACCATCT	TGATGACACC
	22861	GGCGACACCC	GCAGCCGCCT	GCAGCATGACC	GATGTTGAC	TTCAACGAAC	CCAGCAGCAG
	22921	CGGAACCTCA	CGCTCCTGCC	CGTACGTGCG	CAGAATCGCG	TGCGCCTCGA	TGGGATCGCC
15	22981	CAGCGTCGTC	CCC GTCCCGT	GGCCTCCAC	CACGTCCACG	TCGGCGGGGG	CGAGCCJCGC
	23041	CTTGTGGAGG	GCCTGGCGA	TGACCGCCTG	CTGGGAGGGG	CCGTTGGGTG	CGGAGATGCC
	23101	GTTGGAGGCG	CCGTCCCTGGT	TGACGGCGGA	GGAGCGGAGC	ACCCGAGGAA	CGGTGTGTCC
	23161	GTTGCGCTCG	GGCGTGGAGA	GCTTTGAC	GACGAGGAGC	CCGGCCCCCT	CGCGAAACC
	23221	GGTGCCGTCC	GGCGCGTCAG	CGAACGCCCT	GCACCGTCCG	TCCGGCGCGA	CGCCGCCCTG
20	23281	CCGGGAGAAC	TCCACGAAGG	TCTGTGGTGA	TGCCATCACT	GTGACACCAC	CGACCAGCGC
	23341	CAGCGAGCAC	TCCCCGGTCC	GCAGCGCCTG	CCC GGCGCTGG	TGCAGCGCGA	CCAGCGACGA
	23401	CGAACACGCC	GTGTCGACCG	TGACCGCCGG	ACCCCTCCATG	CGGAAGAAGT	ACGACAGCCG
	23461	TCCGGCGAGC	ACCGCGGGCT	GTGTCGCTGTA	GGCGCGGAAT	CCGCCCGAGGT	CCGCGCCCGT
	23521	GCCGTAGCCG	TAGTACAAGC	CGCCGACGAA	GACGCCGGTG	TCGCTGCCGC	GCAGGGTGTC
25	23581	CGGCACGATG	CGGGCGTGT	CGAGCGCCTC	CCAGGCGATT	TCGAGGAGGA	TCCGCTGCTG
	23641	CGGGTCGAGT	GGGGTGGCCT	GGCGCGGACT	GATGCCGAAG	AACCGGGCAT	CGAAGTCGGC
	23701	GGCGCCCGCG	AGTGCGCCGG	CCC GCGCGGT	GGGGGACTCG	GGCGCGGGCGT	GCAGCGCCGGC
	23761	CACGTCCCAG	CCCGCGTCGG	TGGGGAAAGTC	GCCGATCGCG	TCGCGGCCGT	CCGCGACGAG
	23821	CTGCCAACAGC	TCTCCCGGT	AGGTGACGTC	GCCC GGCGACT	CGGCAGGCCA	TGCGGACGAC
	23881	GGCGAGCGGC	TCGTTGCCG	GGCGCGCAG	CGCGGTGTT	TCCCGGCCGA	GCTGCGCGTT
30	23941	GTCTTGACC	GACGTCCGCA	GGCCTCGAT	CAGGTGTTTC	TGCGCCATCG	CCTCATCCCT
	24001	TCAGCACGTG	CGCGATGAGC	GGCTCTCGT	CCATGTCGTC	GAACAGTTCG	TCGTCCGGCT
	24061	CCCGCGTCGT	GGTGTCTCGC	GGTGCCTGTC	CCGGTGGTT	ACCGCCGTCC	GGGGTCCCGT
	24121	TGTCGTCCGG	GGTCCCGTTG	ACGTCGGGGG	CCAGGAGGGT	CAGCAGATGA	CGGGTGAGCG
35	24181	CGCCGGCGGC	GGGATAGTCG	AAGACGAGCG	TGGCCGGCAG	CGGAATGCCG	AGGGCCCTCGG
	24241	AGACCCGGTT	GGCGAGGCCG	AGCGCGGTGA	GCGAGTCGAC	CCCCAGGTCC	TTGAACGCCG
	24301	TGGTGGCCGT	GAACGCCGCC	GGCTCGGTGT	GGCCCAGCAG	GGTGGCGGCCG	GTGTCGCCGA
	24361	CGACGCCGAG	CAGCACCTGT	TCCCGTTCC	TGTGGGGCAG	GTCCGGCAGG	CGTTCCAGCA
	24421	GGGAGCCGCC	GTGGTGCAGG	GAGGCCGGG	TGGGGCGCTG	GATGGTGCAGC	CACAGCGGTG
40	24481	ACGGGTCGCC	GGGCCCGGGT	GGGGCGGTG	CCACGACAC	GGCTCCCG	GTGGCGCACG
	24541	CGGCGTCGAG	GAGGTGGTC	AGCCGGTCCG	CCGCGGGCGT	GAACGCCACG	GCCGGCAGGC
	24601	CTTGTGCCCG	GGCAGGGTCG	GCCAGGGCCT	GGAGCGGTCC	GGCCGCCCTCG	CCGGACGGAA
	24661	CGCGAGAAC	GAACCGGGTC	AGGTCGAGGT	CGCGGGTCAG	GGGGTGCAGT	TCCCAGGCCG
	24721	ACTCGCGGT	GGCGTCCGCG	TGGACGACCG	CGGTACCCG	GGTTCCGGC	ACTGTGCCCG
45	24781	GCTCGTACCG	GATCACTTCG	GGCGCGTGT	CGCCGAGGTG	TCCCGCGAGT	TCCCTCGAAC
	24841	CGCCCGCGAG	GAGGACGGT	TCGCGTACG	AGGCCGCCG	CGTGGTGGGC	CGGGCGGGGA
	24901	CGAGGCCGGG	CGCTTCGAGG	CGCCCGTCGG	CCAGGCGCAG	GTGCGGTTCG	TCGAGGCCGG
	24961	AGAGGCCGGC	GGCGCGGCCG	GGGGTGACCG	TGTCGGTGGT	CTCCACGAGC	ACGAGCCGGC
	25021	CCGGTCCCGC	GGTGTGAGC	AGTGCAGGCA	GGCACCCGC	GACGGGCCCG	GCCTCGCGG
	25081	ACACCACCA	CGTGGCGCCG	GGGGTCCCTG	GGTCGTCCAG	TGCCGTACGG	ACCTCGTCGG
50	25141	GACCGATAAC	GGGGACGACG	ATGACGTGCG	GGCGTGGCGTC	TGCCCGAGG	TCGGTGTACC
	25201	GGCGGGCCGT	GGTGGCGGGT	GGCGCCGGGG	CCCGGACGCC	GGTCCAGGTG	CGCCGGAACAA
	25261	GCCGCACTGC	CCC GTCCGGG	CCCGTCGTGG	GGGGGGCCG	GGTGTATGAGC	GAGCCGATCT
	25321	GAGCCACCGG	CCGTCCCAGT	TGTCGGCGA	GGTGCACCGC	GGCCGCCGCC	TGCCCCCTCGC
	25381	CGTGGACGAA	GGTGCAGGCC	AGTTCTGTTG	GGCGCGTGGT	GTGGACACGG	ACGCCGGTGA
55	25441	ACGCCAACCG	CAACCGTAC	GGCGCGTTCT	GGGGCGCCCG	GCCGATGCTG	CCCGCTTGCA
	25501	GGCGGGCCGT	GGTGGCGGGT	GGCGCCGGGG	CCCGGACGCC	GGTCCAGGTG	CGCCGGAACAA
	25561	CGTCGAGGCC	GACTTCGGCG	CAGACGGTGT	CTCCGTGGCT	CCACGCCCG	GACATGCCGC
	25621	GGAACCTCGG	GGCGAACTCG	TATCCCGCTG	CGTCGAGTCG	CTGGTAGAAG	GCCGCCGACGT
	25681	CGACCGGTTC	GGCGTGCTCG	GGCGGCCAGG	GGGGCGCCGT	GGTGGCCGGT	TCGGTGGTGG
60	25741	CGATGCCGGC	GAAGGCCGGAG	GGGTGGCGGG	TCCATGTCG	GTGCCGTCC	GTCCGGCGT
	25801	GGACGCCGAC	GGCAAGGCCGT	CCGGTGTGCT	GGGGCGCCGG	GACGGTCACG	CGCACCTGGA
	25861	CGGEGCCGGT	GGCAGGGCAGG	ACCAGCGGTG	TCTCGACGAC	CAGTTGTCG	AGCAGGTGCG
	25921	AGCCTGCCCTC	GTGGCGGCCG	GGTCCGGCCA	ATTCCAGGAA	GGCGGGTCCG	GGCAGCAGTA
	25981	GGCGGCCGTC	GACGGAGTGA	CCGGCCAGCC	ATGGGTGGGT	GGCAGCGAG	AACCGGCCGG

26041	TGAGCAGCAC	CTCGTCGGAG	TCGGGGAGCG	CCACCGACGC	GGCGAGCAGC	GGGTGGTCGA	
26101	CGCGCTCGAG	TCCGAGGCCG	GAAGCGTCCG	TGCCGCCCGC	GGTCTCGATC	CAGTAGCGCT	
26161	CATGGTGGAA	GGCGTATGTG	GGCAGGTCGT	GTGCCGTCGC	CGTCGCGGGG	ACGACGCCG	
26221	CCCAGTCGAC	GGGCACGCCG	GTTGTGTGCG	CCTCGGCCAG	CGCGGTGAGC	AGCCGGTGGA	
5	26281	CTCCCCCGCC	GGGGCGGAGC	GTGGCGACGG	TCGCGCCGTC	GATCGCGGGC	AGCAGCACGG
26341	GGTGCGCGCT	GACCTCGACG	AACACGGTGT	CACCCGGCTC	GCGGGCAGCG	GTCACGCCG	
26401	TGGCGAAGCC	TACGGGGTGG	CGCATGTTGC	GGAAACCAGTA	CTCGTCGTCG	AGCGGCCGCGT	
26461	CGATCCAGCG	TTCGTCGGCG	GTGGAGAAC	ACGGGATCTC	GGCGTGCACG	GAGGTGGTGT	
10	26521	CGCGACGAT	CCGCTGGAGT	TCGTCGTACA	CGGGGTGCGAC	GAACGGGGTG	TGGGTGGGCG
26581	AGTCGACGGC	GATGCGGCCG	ACCCAGACGC	CGCGGGCCTC	GTAGTCGGCG	ATCAGCGTTT	
26641	CGACGGCGTC	CGGGCGCCCG	GCGACGGTCG	TGGTGGTGGC	GGCGTTGCGG	CCCGCGACCC	
26701	AGACGCCGTC	GATCCGGGCG	GCATCCGCCCT	CGACGTGCGC	GGCCGGGAGC	GCGACCGAGC	
26761	CCATCGCGCC	GGCTCCGGCG	AGTTGCGCAG	GGAGCAGGAG	AACGCTGCGC	AGCGCGACGA	
15	26821	GGCGGGCACC	GTCCCTCAGG	GTGAGCGCTC	CGGCGACACA	GGCCGCGGCG	ATCTCGJCC
26881	GGGAGTGTCC	GATGACCGGC	TCCGGGCGTA	CGCCCGCGGC	CTCCCACACG	GC GGCCAGCG	
26941	ACACCATGAC	GGCCCAGCAG	ACGGGGTGCA	CGACGTGAC	GC GGCGGGTC	ACCTCCGGT	
27001	CGTCGAGCAT	GGCGATGGGG	TCCCAGCCCG	TGTGCGGGAT	CAGCGCGTCG	GCGCATTTGC	
27061	GCATCCTGGC	GGCGAACACC	GGGGAGGCGG	CCATCAGTC	GACGCCCATG	CCCGGCCACT	
20	27121	GGGGTCTTG	TCCGGGGAAAG	ACGAAGACGG	TGCGGGCTC	GGT GAGCGCC	GTGCGGTGTA
27181	CGACGTCGTC	GTCGAGCAGC	ACGGCGCGGT	GGGGAACGT	CGTACGCC	GCGAGCAGGC	
27241	CCGCGCGCAT	GGCGCGCGG	TCG TGGCGG	ACGGGGCGGC	GAGGTGCTCG	CGGAGTCGGC	
27301	GGACCTGGCC	GTCGAGGGCC	GTGGCGGTCC	GGCGAGAC	GGG CAGTGGT	GTGAGCGGCG	
27361	TGGCGATCAG	CGGCTCACCG	GGCTCGAGG	CCGACGGCTC	CTCGGCCGGC	GGCTCCCCGG	
25	27421	CCGGGTGGC	TTCCAGCAGG	ACGTGGCGT	TGGTGGCGCT	GACGCCGAAG	GAGGACACAC
27481	CGGCGCGCCG	CGGGCGGTG	GTCTCGGGC	AGGGCGGGC	ATCGGTGAGG	AGTCGACGG	
27541	CGCCGCCGT	CCAGTCGACG	TGCGAGGAGC	GGGTGTCCAC	GTGCAAGGGT	CGCGGCAGGG	
27601	TGCCGTGCCG	CATGGCGAGG	ACCATCTTGA	TGACACCGGC	GACACCCGCG	CGGGCCTGAG	
27661	TGTGGCCGAT	GT TGGACTTC	AGCGAGCCA	GCAGCACCG	GGTGTGCGC	CCCTGCCGT	
30	27721	AGGTGCCAG	CACCGCCTGT	GCCTCGATGG	GATCGCCAG	CCTGGTGCGC	GTGCCGTGCG
27781	CCTCACCGC	GTCCACGTC	GCCGGGGTGA	GCCC CGCGTT	GGCCAGGGCC	TGCCGGATCA	
27841	CCCGCTCCTG	CGAGGGCCCG	TTCGCGCCG	ACAACCGTT	GGAAAGCACCG	TCC TGGTTGA	
27901	CCGCCAAC	CCGGACAAAC	GCCAGCACAC	GGTGGCGTT	GGCGCTCGCA	TCGGAGAGCC	
27961	TCTCGACGAT	CAGCACACCG	GACCCCTCGG	CGAAACCGGT	GGCGTCA GCGC	GCATCCJCGA	
28021	ACGCCCTGCA	GGCGCGTCG	GGCGCGAGAC	CCC GCTGCTG	GGAGAACTCG	ACGAAGCCGG	
35	28081	ACGGCGAGGC	CATCACCGTG	ACGCCGCCGA	CCAGGGCGAG	CGAGCATTG	CCGGAGCGCA
28141	GTGACTGCC	GGCCTGGTGC	AGC GCCACCA	GGCGACGAGA	ACACGCCGTG	TCGACCGTGA	
28201	CCGCCGGACC	CTCCAGACCG	TAGAAGTACG	ACAGCCGACC	GGACAGCACA	CTGGTCTGGG	
28261	TGCCGGTCG	GGCGAAACCG	CCCAGGTGCG	TGCCGAGTCC	GTACCCGTG	GAGAAGGC	
28321	CCATGAACAC	GGCGGTGTCG	CTTCCCGCGA	GGCACTCCCG	GAGGATCCC	GGGTGTCCA	
40	28381	GCGCCTCCCA	CGAGGTCTCC	AGGACCAGAC	GCTGCTGCGG	GTCCATCGCC	AGCGCTCAC
28441	GCGGACTGAT	CCCGAAGAAC	GCCCGTGC	AGTCCGCCAC	CCCCCGAGG	AAGCCACCAT	
28501	GACGCACGGT	CGACGTGCC	GGATGATCCG	GATCGGGATC	GTACAGCCCG	TCCACGTCCC	
28561	AACCACGGTC	CGTCGGAAAC	GGCGTGA	CGTCACCA	CGACTCCAGC	AGCCGCCACA	
28621	AGTCCTCCGG	CGACCGCAGC	CCACCCGGCA	GCCGGCAGGC	CATCCCCACG	ATCGCAAACG	
45	28681	GCTCGCCTG	CCGGACGGCC	GCGGTGCG	TGCGGGTCCG	CGATGCCGTC	CGGCCGAC
28741	GCGCCGCGGT	GAGCTCGCC	GCGACGGCGC	GGGGCGTCG	GAAGTCGAAG	ACCGCGTGG	
28801	CGGGCAGCCG	TACGCCGTC	GCCTCGGTGA	AGGC GTTGCG	CAGCCGGATC	CCATGAGCG	
28861	AGTCGACGCC	GAGTTCTTG	AACGTGGCG	TCGCCCTGAC	CCGTGCGGCA	CCGTGCG	
28921	CGAGTACGGC	CGCGGTGCA	TGCGGACGA	CGGCGAGCAC	GTCTTTTCG	GGTCCCGCG	
50	28981	CGGAGAGCCG	CGCGATCCCG	TCGGCGAGGG	TGGTGGCGCC	GGCCGCCCGG	CGCCGCGGCT
29041	CCCGCGCGG	TGCGCGCAGC	AGGGCGAGC	TGCCGAGGCC	GGCCGGGTG	GGCGCGACCA	
29101	GCGCCGGTC	CGAGGACCGC	AACGCCGCT	CGAACAGCGT	CAGTCCGCGT	TCGGCGTCA	
29161	GCGCCGTCAC	GGCGTCGCG	CGCATGCGG	CGCCGGTGCC	GACCGTCAGC	CCGCTCTCCG	
29221	GTTCCACAG	GGCCCAGGGC	ACGGACAA	CGGGCAGTCC	GGCTGCCCGG	CGCTGTTCGG	
55	29281	CCAGCGCGTC	GAGGAACCGC	TTCGCGGCCG	CGTAGTTG	CTGTCGGGG	CTGCCGAGCA
29341	CACCGCGCC	CGACGAGTAG	AGGACGAACG	CGGCCAGTTC	CGTGTCTGG	GTGAGTCGT	
29401	GCAGGTGCCA	CGCGCGTCC	ACCTTCGGC	GCAGCACCGT	CTCGAGGCCG	TCGGGGGTGA	
29461	GCGCGGTGAG	GACGCCGTCG	TGAGGACGG	CCGCGGTG	CACGACGGCC	GTGAGCGGGT	
29521	GCGCCGGTC	GATCCCCGCC	AGTACGGAGG	CGAGTTCGTC	CCGTCGGCG	ACGTCCGAGG	
60	29581	CGATCGCCGT	GACCTCGGCC	CCGGGCACGT	CGCTCGCCG	GGCCGTCGCG	GACAGCATCA
29641	GCAGCGCGC	CACGCCGTG	CGTTCGACG	GGTGGCGC	GATGATGCC	GCCAGCGTCC	
29701	CGGAGCCACC	GGT GACGAGC	ACGGTGCCTG	CCGGGTG	GGCCGGAGCG	TCACCCGCCG	
29761	GGACCGCCGG	GGCCAGACGG	CGGGCGTACA	CCTGGCGTC	ACCGAGCACC	ACCTGGGGCT	
29821	CATCGAGCGC	GGTGGCCGCT	GCGAGCAGCG	GCTCGGGCGT	GTCCGGGGCG	GGGTGCGACGA	

29881 GGACGATCCG GCCGGGGTGT TCGGCCTGCG CGGTCCGCAC CAGTCCGGCG GCCGCGGCCG
 29941 ACGCGAGACC GGGCCCGGTG TGGACGGCCA GGACCGCGTC GGCGTACCGG TCGTCGGTGA
 30001 GGAAGCGCTG CACGGCGGTG AGGACGCCGG CGCCCAGTTC GCGGGTGTG TCGAGCGGGG
 30061 CACCCCGGCC GCCGTGCGCG GGGAGGATCA CCACGTCCGG GACCGTCGGG TCGTCGAGGC
 5 30121 GGCGGGTCGT CGCGGTGCGT GGCAGCAGCT CCGGGAGCTC GGCCAGCAC GGGCGCAGCA
 30181 GGGCCGGAAC GGCTCCCGTG ATCGTCAGGG GGCACCGTGC CACGGCGCCG ATGGTGGCGA
 30241 CGGGCCCGCC GGTCTCGTCC GCGAGGTGTA CGCCGTCAGC GGTGACGGCG ACACGTACCG
 30301 CCGTGGCGCC GGTGGCGTGG ACGCGGACGT CGTCGAACGC GTACGGAAGG TGGTCCCCTT
 30361 CCGCGGCGAG GCGGAGTGC GCGCCGAGCA GCGCCGGGTG CAGGCCGTAC CGTCGGCGT
 10 30421 CGGCGAGCTG TCCGTCGGCG AGGGCCACTT CCGCCCAGAC GGCACCGTGC TCGGCCCAGA
 30481 CGGCGCGCG GCGGGGCAGC GCGGGCCCGT CCGTGTACCC GGCTCGGGCC AGACGGTCGG
 30541 CGATGTCGTC GGGGTCCACC GGCGGGGCCG TGGCGGGCGG CCACGTCGAC GGCATCTCCC
 30601 GCACGGCCGG GGCGTCCCGC GGGTCGGGGG CGAGGATTCC GTGCGCGTGC TCGGTCCACT
 15 30661 CCCCCGCCGC GTGCCGCGTG TGCACGGTGA CCGCGCGGCCG GCCGTCGCCGCC CCGGCGGCC
 30721 TCACCGTGAC GGAGAGCGCG AGCGCACCGG ACCGCGGCAG CGTGAGGGGG GTGTCCACGG
 30781 TGAACGTGTC GAGGGCGCCG CAGCCGGCTT CGTCGCCCCG CCGGATCGCC AGATCCAGGA
 30841 GGGCCGCGGC GGGCAGCAC C GCGAGGCCGT GCAGGGAGTG CGCCAGCGGA TCGGCGCGT
 30901 CGACCCGGCC GGTGAGCACC AGGTGCGCCG TGCCGGGAG GGTGACCGCC GCGGTCAGCG
 20 30961 CGGGGTGCGC GACCGGGCGT TGTCGGGCCG GGGCCGCGTC GCCCGCGGCC TGGGTGCGA
 31021 GCCAGTAGCG GACCCGCTCG AACGGGTAC TCGCGGGTGC CGAGGCGCGT GCGGGCGCG
 31081 GGTGATGAC CTTCGGCCAG TCGACCGTGA CGCCGTCGGT GTGCAAGCGG GCGAGCGCG
 31141 TCAGGGCGGA TCGCGGTTCG TCGTCGGCGT CGACGATCGG GATGCCGTCG ACGAGTCGGG
 31201 TCAGGCTCCG GTCCGGGCCG ATCTCCAGGA GCACCGCCCG GTCGTCGCGC GCGACCTGTT
 25 31261 CCCCCAACCG GACGGTGTGC CGGACCGTGC GTACCCAGTA CTCCGGCGTG GTGCAGCGG
 31321 CGCCCGCGGC CATCGGGATC CTCGGCTCGT GGTACGTCA GCTCTCCCGC ACCTTGCGGA
 31381 ACTCCTCGAG CATCGGCTCC ATCCCGCGCG AGTGGAACCG GTGGCTGGTC CGCAGGGCGG
 31441 TGAAGCGGCC GAGCCGGGCC GCGACGTCA GCACCGCCCTC CTCGTACCCG GAGAGCACGA
 31501 TCGACGGCGG CCCGTTGACC GCGGCGATCT CCACGCGTC CGCAGCGAGC GGCAGCGCGT
 31561 CCCGTTCCGA CGCGATCACG CGGGCCATCG CCCCAGCGGA CGCAGCGGCC TGCACTAGGC
 30 31621 GGGCCCGTGC GGACACCAGC CTGACCGCGT CCTCCAGGGG CGACAGCGCC GCGACGTACG
 31681 CGGCGGCCAG CTCGCCGATC GAATGGCCA CGAAGGGCGTC CGGGCGTACG CCCCACGCC
 31741 CGAGCTGTGC GCGAGGTGCG ACCTGGAGCG CGAACACCGC GGGCTGGCG TAACCGGTGT
 31801 CGTGGAGGTC GAGCCCGGGC GGCACGTCA GGGCGTCCAG CACCTCGCGG CGAGTGC
 31861 CGAAGACGTC GTAGGGCGCG GCCAGTCCGT CGCCCATGCC GGGACGTTGT GAGCCCAGTC
 35 31921 CGGAGAAAGAG CCACACGAGG CGGCGGTCCG GTTCTGCGGC GCGGGTGACC GTGTCGGTGC
 31981 CGATCAGCGC GGCCCGGTGC GGGAAAGGCCG TGCGGGCGAG CAGGGCGCGC GCAACCGCG
 32041 GCTCGTCTC CTCGCCGGTG GCGAGGTGGG CGCGCAGGGC GTGTACCTGT GCGTCGAGTG
 32101 CCTGCGGGGT GCGTGCCAG AGCAGCAGGG CAAGCGGTCC GGTGTCGGGT GCGGGGGCGG
 32161 GTTCGGGGGC CGGTGGGGGG TGGCTTCGA GGATGATGTG AGCGTTGGTG CCGCTAACGC
 40 32221 CGAAGGAGGA CACCCCGGGC CGCCGTGGGC GGTGGTTTC GGGCCAGGGG CGGCGTCTCG
 32281 TGAGGAGTTG GACGGCGCCG GCGTCCAGT CGACGTGCGA GGACGGCGTG TCCACGTGCA
 32341 GGGTGCAGCG CAGGGTGCCG TGCCGATGG CGAGGACCAT CTTGATGACA CGGGCGACCG
 32401 CCGCGGCCGC CTGAGTGTGG CGATGTTGG ACTTCAGCGA GCCCAGCAGC ACCGGGGTGT
 32461 CGCGATGCTG CCCGTAGGTG GCCAGTACCG CCTGCGCTC GATGGGGTGC CCCAGCCTGG
 45 32521 TCCCGGTGCC ATGCGCTCG ACAGCGTCCA CATCGGCCGG GGTGAGCCCG CGTGGTGGCCA
 32581 GCGCCTGCCG GATCACCCGC TCCTCGACG GCGCGTTCCG CGCCGACAAC CGTGGTGGAA
 32641 CACCGTCTG GTGACCGGCC GAAACACGCA CGACCGCCAG GACATTGTGG CCGTGCCT
 32701 CGGCGTCGGA GAGCCTCTCG ACGATCAGCA CACCGGATCC CTGGCGAAA CGGTGCCCCAT
 32761 CAGCCCCATC CGCGAACGCC TTGCAGCGGC CGTCCGGGG AAGGGCCCCGC TGCTGGGAGA
 50 32821 AGTCCACGAA GCGGACGGC GAGGCCATCA CGTGCACGCC CGCGAACACAG CGAGCGAGC
 32881 ACTCCCCCGA CGCGACGCGAC TGCCCCGGCTA GGTGAGCGGC CACCAAGCGAC GACGAACACG
 32941 CCGTGTCCAC CGTGACCGCC GGACCCCTCA AACCGTAGAA GTACGACAGC CGACCGGACA
 33001 GCACACTGGT CTGGGTGCTG GTGGCACCGA AACCGCCCGC GTCGGCTCCA GTGCCGAGAC
 33061 CGTAGAAGTA GCGCCCATG AACACGCCGG TGTCGCTTCC GCGCAGCGAC TCCGGGAGGA
 55 33121 TCCCGGTGT TTCCAGCGCC TCCCAAGGAG TCTCCAGGAC CAGACGCTGC TGGGGTCCA
 33181 TCGCCAGCGC CTCACCGGAGA CTGATCCCGA AGAACGCCG GTCGAAGTCC GCCACCCCGG
 33241 CGAGGAAGCC ACCATGACGC ACGGTCGACG TGCCCGGATG ATCCGGATCG GGATCGTACA
 33301 GCGCGTCCAC GTCCCCAACCA CGGTCCGTG GAAACGCCGT GATCCCGTCA CCACCCGACT
 33361 CCAGCAGCGC CCACAAGTCC TCCGGCGACG CGACCCCAAC CGGCAGCGCG CAGGCCATCC
 60 33421 CCACGATCGC CAACGGCTCG TCCTGCGGGA CGGCGCGGGT CGGGGTACGC CGCCGGGTGG
 33481 TGGCCCGCGC GCGGGCCAGT TCGTCCAGGT GGGCGCGAG CGCCTGCGCC GTGGGGTGGT
 33541 CGAAGACGAG CGTAGCGGGC AGCGTCAGGC CGTCGCGTC GGCCAGCGGG TTGCGCAGTT
 33601 CGACGCCGGT CAGCGAGTCG AAGCCACTT CCCTGAACGC CGCGCGGGGT GCGATGGCGT
 33661 GGGCGTCGCG GTCGGCCGAG ACCCGGGCAG CGCTGGTACG GACGAGGTG AGCATGTCG

	33721	GCGCGGCCGG	AGGTGCGGAC	GTGCGCCGGA	CGGCCGGCAC	GAGGGTGCCT	AGGACCGGGC
	33781	GGACCCGGTC	GGACGCGGCCG	ACGGCGGCGA	GGTCGAGCCG	GATCGGCACG	AGCGCGGGCC
5	33841	GGTCGGTGTG	CAGGGCGCGG	TCGAACAGGG	CGAGCCCCG	TGCGGCCGTC	ATCGGGGTCA
	33901	TGCCGTGCG	GGCGATGCGG	GCCAGGTCTGG	TGGCGGTCTAG	CCGCCCGCCC	ATCCCGTCCG
	33961	CCGCGTCCC	CAGTCCCCAG	GCGAGCGAGA	CGGCGGGCAG	CCCCCTGGTGG	TGCCGGTGGC
	34021	GGGCGAGCGC	GTCGAGGAAC	GCGTTGCCGG	TCGCGTAGTT	GGCCTGACCC	GCGCCGCGGA
	34081	ACGTGGCGGA	TATGGACGAG	TACAGGACGA	ACGCGGCCAG	GTCGAGATCG	CGCGTCAGCT
	34141	CGTGCAGGTG	CCAGGCAGACG	TCCGCCTTGA	CCCAGCAGAC	GGCGTCCCAC	TGCTCCGGCC
10	34201	GCATGGTCGT	CACGGCCGCG	TCGTCGACGA	TCCCAGGCCAT	GTGCACGACG	GCGCGCAGCC
	34261	GCTGGGCGAC	GTCGGCGACG	ACTGCGGCCA	GTCGTCGCG	GTCGACGACG	TCGGCGGCCA
	34321	CGTACCGCAC	GCAGGTGTC	TCCGGCGTGT	CGCCGGGCCG	GCCGTTGCGG	GACACCACGA
	34381	CGACCTCGGC	GGCCTCGTGC	ACGGTGAGCA	GGTGGTCCAC	GAGGAGGCAG	CCGAGCCCCG
	34441	CGGTGCCGCC	GGTGACGAGG	ACGGTCCCAG	CGGTCAGCAG	GGAGGTTCCG	GTGGCCGCCG
15	34501	CGACACGGCG	CAGACGGGCC	GCACGCGCTG	TGCCGTGCCG	GACCCGGACG	TGCGGCTCGT
	34561	CGCCGGCGGC	GAGCCCGGCC	GCTATGGCGG	CGGGCGTGAT	CTCGTCCGCT	TCGATCAGGG
	34621	CGACGCGGCC	GGGATGCTCC	GTCTCCGCG	TCCGGACAG	GCCGCCGAGC	GCTTCCAGCG
	34681	CGGGATCGCC	GGTACGGGTG	GCCACGATGA	GCCGGGATCG	CGCCCGAGCG	GGCTCGGCCA
	34741	GCCAGGTCTG	CACGGTGGTG	AGCAGGTCTG	GGCCCGAGCTC	CCGGGTCCGG	GCGCCGGGCC
20	34801	AGGTGCCCGG	GTCGCCGGGT	TCCACGGCCA	GGACCACGAC	CGGGGGGTGC	TCGCCGTCCG
	34861	GCACGTGCGC	GAGGTACGTC	CAGTCGGGG	CGGGTGACGC	GGGCACGGGC	ACCCAGGCCA
	34921	TCTCGAACAG	CGCCTCGGC	TCGGGGTCGG	CGGCCCCGAC	GGTCAGGCTG	TCGACGTCAA
	34981	GGACCGGTGA	GCCGTGCTCG	TCCGTGGCGA	CGATGCGGAC	CATGTCGGGG	CCGACGCGTT
	35041	CCAGCAGCAC	GCCGAGCGCG	GTCGCGGCCG	CGCGCGTGGAT	CCTCACGCCG	GACCAGGAGA
25	35101	ACGCCAGCCG	GCGCCGCTCC	GGGTCCGTGA	AGACCGTCCC	GAGGGCGTGC	AGGGCCCGT
	35161	CGAGCAGCAC	GGGGTGCAGC	CCGTACCGGG	CGTCGGTGAG	CTGTCGGCG	AGGCGGACCG
	35221	ACGCGTAGGC	GCGGCCCTCC	CCCGTCCACA	TCCGGTCTCAT	GGCCCGGAAC	GCGGGCCCGT
	35281	ACGAGAGCGG	CAGCGCGTCG	TAGAAAGCCGG	TCAGGTCGGC	CGGGTCGGCG	TCGGCGGGCG
	35341	GCCAGTCCAC	GGGCTCCGCC	GGACCGCCAG	TGTCCACGCT	CAGCGCTCCG	GTCGCACGTGA
30	35401	GCGCCCAGGG	GCCCGTCCCG	GTACGGCTGT	GCAGACTCAC	CGACCGCCGT	CGGGACACCT
	35461	CGGTTCCGAC	GGTGGCCTGG	ATTCCTGTG	CGCCGTCGCC	GTCGACCAACC	ACCGGCGCGA
	35521	CGATGGTCAG	CTCCCGCATC	TCCGGCGTGC	CGAGCCGGGC	TCCCGCTTCG	GCGAGGAGTT
	35581	CCACGAGCGC	CGAGCCGGGC	ACGATGACCC	GGCCGTCCAC	CTCGTGGTCG	GCGAGCCAGG
	35641	GCTGACGGCG	TACCGAGACA	CCGCGGTGGC	CAGCGCCGCC	TCGCGTCCG	GCGAGGTCTGA
35	35701	CCCACGAGCC	GAGCAGCGGG	TGGCCGGACG	TTCCCGCCGG	TTCCCGTCTG	ATCCAGTACC
	35761	GGTCACCGCG	GAACGGGTAC	GTGGGCGACG	GCACCAACCG	ACCGCTCGCG	AACGACAGG
	35821	TGACGGGCAC	GCCCCGGACC	CAGAGCGCGG	CGAGCGACCG	AGTGAAGCGG	TCCAGGCCG
	35881	CCTCGCCTCG	CCGCACTGTG	CCGGTGACGA	CCGTATGCCG	ATGCCCGGCC	AGCGTGTCT
	35941	CCAGTGGCTG	GGTGAGCAGC	GGATGCGCAG	TGACCTCGAC	GAACCGCGGG	TATCCGCGGT
40	36001	CCGCCAGGTG	GCCGGTCCGCG	GCAGCGAAC	GAACGGTGC	GCGCAAGGTTG	TCGTACCAAGT
	36061	AGGCGCGTC	CGCGGGCCGG	TCCAGCCACG	CCTCGTCCAC	GGTGGAGAAG	AACGGGACGT
	36121	CCGGCGTGC	CGGAGTGTATG	CCGGCGAGAG	CGTCGAGCAG	CGCCGCCGCG	ATCGTTTCGA
	36181	CATGCGCGGT	GTGCGACGCG	TAGTCGACGG	CGATCCGGCG	GGCGCGGGGG	GTGGCGGCCA
	36241	GCAGCTCCCTC	CACGGCGTCG	GCCGACCCGG	CGACAAACGAT	CGACCGGGGT	CCGTTGACCG
45	36301	CGGCGACCTC	CAGGCGCCCG	GCCCACACGG	CGCGCGTCGA	GTCGGCGGGC	GGCACCGAGA
	36361	CCATGCGGCC	CTGCCCCGGC	AGTTGGTGG	CGACGAGTCG	GCTGCGCACC	GCGACGACCT
	36421	TCGCGCGTC	GTCCAGGGTG	AGCACCCCGG	CGACCGCAGGC	CGCGCGGACT	TCGCCCTGGG
	36481	AGTGGCCGAC	GACCGCGGGC	GGGGCGACCC	CGTGCACAGC	CCACAGCTCC	GCCAGCGCCA
	36541	CCATCACCGC	GAACGACGCG	GGCTGCACGA	CATCGACCCG	GTCGAACGCG	GGCGCTCCGG
50	36601	GCCGCTGGGC	GATGACGTCC	AGCAGGTCCC	ATCCGGTGTG	CGGGCGAGC	GCGTGGCGC
	36661	ACTCGCGGAG	CCGCGGGGCC	AACACGGGCT	CGGTGGCGAG	CAGTCGGCA	CCCATGCCGG
	36721	CCCACTGGGA	GCCCTGCCCC	GGGAACCGCA	ACACGACACG	TGTGTCGGTG	ACGTCGGGCG
	36781	TTCCCGTCAC	GGCCCCCGGC	ACTTCGGCAC	CACGGGCGAA	CGCCTCCGCC	TCTCGGGCCG
	36841	GCACGACCGC	CCGGTGGCGC	ATGGCCGTCC	GGGTGGTGGC	GAGCGAGTGG	CCGACCGCGG
55	36901	CCGCGGGCGC	AGTGAGCGGG	GCCAGCTGTC	CCGCGACGTC	CCGCACTCCC	TCCGGGGTCC
	36961	GGGCCGACAT	CGGCCAGAC	ACGTCTCGG	GCACCGGCTC	GGCTCGGGGT	GCGGACACGG
	37021	GTGCGGGCGC	GGCGGGGGGC	CCGGCCTCCA	GGACGACATG	GGCCTTGGTG	CCGCTGATGC
	37081	CGAACGACGA	GACACCCGCA	CGCCGGCGC	GGCCGGTGAC	CGGCCACGGC	TCACTGCGGT
	37141	GCAGCAGCG	GATGTCGCCC	TCCCAGTCGA	CGTGCACGGGA	CGGCTCGTCG	ACGTGCAGCG
60	37201	TGCGCGGCAG	GACGCCGTG	CGCATCGCCA	TGACCATCTT	GATGACGCGC	GCGACGCCG
	37261	CCGCGGGCTG	GGTGTGGCCG	ATGTTGACT	TGAGCGAGCC	GATCAGCAGC	GGATGACACG
	37321	GTTCGCGCCC	GTAGGCCACT	TGCAGGGCCT	GGGCCTCGAC	GGGGTCGCGC	AGACGGGTGC
	37381	CGGTGCCGTG	TGCCTCCACG	GCGTCGACGT	CACCCGGGCC	CAGGCCAGCG	TCGGCGAGCG
	37441	CACGCTGGAT	GACGCCGTG	TGCGCAGGCC	CGTTCGGGGC	GGACAGCCCG	TTCGACGCCG
	37501	CGTCGGAGTT	GACCGCGGAG	CCGCGCACCA	GCGCCAGCAC	GGGGTGGCGC	TGGCGGGTGG

37561	CGTCGGAGAG	CCGCTCCAGC	ACCAAGGACAC	CGGGGCCCTC	GGCGAAGCTC	GTGCCGTCCG	
37621	CGGTGTCCGC	GAAGGCCTTG	GCACGGCCGT	CGGGGGCGAG	CCCGCGCTGC	CGGGAGAACT	
37681	CGACGAACCC	GTCGTCGTC	GCCATCACCG	TGACACCGCC	GACCAGGGCG	AGCGAGCACT	
37741	CCCCCGAGCG	CAGCGACCGC	CGGGCCTGGT	GCAGCGCCAC	CAGCGACGAC	GAACACGCCG	
5	37801	TGTCGACGGT	GACCGACGGG	CCCTCCAGAC	CGAAGTAGTA	CGAGAGCCG	CCGGAGAGAA
	37861	CGCTGGTCGG	CGTGCCTGGC	GCCCCGAAAC	CGCCCAGGTC	CACGCCCGCG	CCGTAGCCCT
	37921	GGGTGAACGC	GCCCCATGAAT	ACGCCGGTGT	CGCTGCCCG	GACGCCCTTCG	GGCAGGATGC
	37981	CCGCTCGTTC	GAACGCCCTC	CACGACGCTT	CGAGGACCAAG	ACGCTGCTGC	GGGTCCATCG
10	38041	CCAGGCCCTC	ACGCGGGCTG	ATCCCGAAGA	ACGCGGCCGT	GAAGTCGGCG	GCGCCGGTGA
	38101	GGAAGCCGCC	GTGACGCACG	GAAACCTTGC	CGACCGCGTC	GGGGTTCGGG	TCGTAGAGCG
	38161	CGCGGAGGTC	CCAGCCGCGG	TCGGGGGGGA	ACTCGGTGAT	CGCGTCCCCG	CCGGAGTCGA
	38221	CCAGCCGCCA	CAGGTCCCTC	GGTGACCGCA	CGCCACCGGG	CATCCGGCAC	GCCATGGCCA
	38281	CGATCGCCAG	CGGCTCGTTC	CCCGCCACCG	TCGGTGCGGG	CACTGTGCGC	GCGGAGCGG
15	38341	CAGGGGCCGG	CTCACCCCCC	CGTTCTCAT	CCAGGGGGC	GGCGAGCGCG	GCGGGTGTGCG
	38401	GGTGGTCGAA	GACGGCCGTC	CGGGAGAGCC	GTACCCCCGT	CGTCTGGCG	AGGCTGTTGC
	38461	GCAACCGGAC	ACCGCTGAGC	GAGTCGATGC	CGAGGTCCTT	GAACGCCGTC	GTGGGCGTGA
	38521	TCTCGGAGGC	GTCGGGCTGG	CCGAGCACGG	CGGCCGTGGC	CGCACACACG	ATGGCCAGCA
	38581	GGTCACGATC	CGGGTCGCGG	TCGCGGTGCG	GGTTGTCCTC	CGCACGGGC	GCGATGCCGC
20	38641	GCTCGTCCG	CTGCCGGACG	GGCTCGGTGG	GAATGCCGC	GACCATGAAC	GGCACGTCG
	38701	CGCGGAGGCT	CGCGTCGATG	AAGTGGGTGC	CCTCGGCCCT	GGTGAGCGGC	CGGAACCCGT
	38761	CGCGCACCCG	CTGCCGGTGTG	CGCTCGTCAA	GTTGTCGGGT	GAGGGTGCTG	GTGGTGTGCC
	38821	ACATGCCCA	GGCGATGGAG	GTGGGGGTT	GGCGGAGGGT	GTGGCGGTGG	GTGGCGAGGG
	38881	CGTCGAGGAA	GGCCTTGGCG	GGCGCTAGT	TTCCCTGTCC	GGGCTGCCG	AGGACGGCGG
25	38941	CGGCCTGGA	GTAGAGGACG	AAGTGGGTGA	GGGGTTGGTT	TTGGGTGAGG	TGGTGCAGGT
	39001	GCCAGGCCG	GTTGGCTTGTG	GGGTGGAGGA	CGGTGGTGAG	GCGGTCGGGG	GTGAGGGCGT
	39061	CGAGGATGCC	GTCGTCGAGG	GTGGCGCGG	TGTGGAAGAC	GGCGGTGAGG	GGTTGGGGGA
	39121	TGTGGCGAG	GGTGGTGGCG	AGTTGGTGGG	GGTCGCCAAC	GTCCGAGGGG	AGGTGGGTGC
	39181	CGGGGGTGGT	GTCGGGGGGT	GGGGTGCGGG	AGAGGAGGTA	GGTGTGGGG	TGGTTCAAGGT
	39241	GGCGGGCGAG	GATGCCGGCG	AGGGTGCSSG	AGCCGCCGGT	GATCATGATG	GCGTGTTCGG
30	39301	GGTTGAGGGG	GGTGGTGGTG	GGTGGGGTGG	TGGTGTGGAG	GGGGGTGAGG	TGGGGTCCGGT
	39361	GGAGGGTGTG	GTGGGTGAGG	CGGAGGTGGG	GGTGGTCCGAG	GGTGGCGAGT	TGGGCCAGGG
	39421	GGAGGGGAGT	GTGGGGGTTG	TCGGTTTCGA	TGAGGGCGGAT	CGGGTGGGG	TGTTCGTTCT
	39481	GGGCGGTGCG	GGTGAGGCCG	GTGACGGTGG	CGCCGGCGGG	GTGGTGGTG	GTGTGGACGA
	39541	TGAGGGTGTG	GTCGGGTGGTG	GTGAGGTGGT	GTTGCAGGGC	GGTCAGGACC	CGGGTGGCGC
35	39601	GGGTGTGGC	GGGGGTGGGT	ATGTCCTCGG	GGTCGTCGGG	GTGGCGGGCG	GTGATCAGGA
	39661	CGTGTCCCTC	GGGCAGGTCA	CCGTCGTAGA	CCGCCTCGGC	GACCGCGAGC	CACTCCAACC
	39721	GGAGCGGGTT	CGGCCCGCAC	GGGGTGTGCG	CCCGCTCCCT	CAGCACCAAGC	GAGTCCACCG
	39781	ACACGACAGG	ACGGCCATCC	GGGTCGGCCA	CGCGCACGGC	GACGCCGGCC	TCCCCCGGG
40	39841	TGAGGGCGAC	GGCAGCCGCG	GGGGCCCCGG	TGGCCTTCAG	GCGCACGCC	GTCCAGGAGA
	39901	ACGGCAGCTC	GATCCCGCCG	CCCGCGTCGA	GGCGCCCGGC	GTGCAGGGCC	GCGTCGAGCA
	39961	GTGCCGGATG	CACACCGAAA	CCGTCGCCCT	CGGCGGCCCTG	CTCGTCGGGC	AGCGCCACCT
	40021	CGGCATACAC	GGTGTACCA	TCACGCCAGG	CAGCCCGCAA	CCCCCTGGAAC	GCCGACCCGT
	40081	ACTCATAACC	GGCATCCCCC	AGTCGTCAT	AGAACCCCCA	GACGTCGACG	GCCGCGGCCG
45	40141	TGGCCGGCGG	CCACTGCGAG	AACGGCTCAC	CGGAAGCGTT	GGAGGTATCC	GGGGTGTCCGG
	40201	GGGTCAAGGT	GGCGCTGGCG	TGCGGGTCC	AGCTGCCGT	GCCCTCGGTA	CGCGCGTGG
	40261	CGGTCAACCG	CCGCCGTCCG	GCCTCATCGG	CCCCCTCCAC	GGTCACCGAC	ACATCCACCG
	40321	CTGCGTCAC	CGGCACCCACG	AGCGGGGATT	CGATGACCG	TTCATCCACC	ACCCCGCAAC
	40381	CGGTCTCGTC	ACCGGGCCCGG	ATGACCAGCT	CCACAAACGC	CGTACCCGGC	AGCAGAACCG
	40441	TGCCCCGCAC	CGCGTGTACA	GCCAGCCAGG	GATGCGTAGC	CAATGAGATC	CGGCCGGTGA
50	40501	GAACAACACC	ACCACGTGCG	TCGGGGGCA	GTGCTGTGAC	GGGGCCAGC	ATCGGATGCG
	40561	CGCCCCGGT	CAGCCCGGCC	CGGGACAGGT	CGGTGGCACC	GGCCGCCTCC	AGCCAGTACC
	40621	GCCTGTGCTC	GAACGCGTAG	GTGGCGAGAT	CCAGCAGCCG	CCCCGGCACC	GGTTCGACCA
	40681	CGTGCCCCCA	GTCCACCCCC	GCACCCAGAG	TCCACGCCGT	CGCCAACGCC	CCCAGCCACC
	40741	GCTCCCAGCC	ACCGTCACCA	GTCCGCAACG	ACGCCACCGT	GGGGCCTGT	TCCATGCCG
55	40801	GCAGCAGCAC	CGGATGGGCA	CTGCACTCCA	CGAACACCGA	CCCCGCCAGC	TCCGCCACCG
	40861	CCGCATCCAG	CGCGACAGGG	CGACCGAGGT	TCCGGTACCA	GTACCCCTCA	TCCACGGCT
	40921	CGGTCAACCA	GGCGCTGTCC	ACGGTCGACC	ACCACGCCAC	CGACCCGGTC	CCGCCGGAAA
	40981	TTCCCTTCAG	TACCTCAGCG	AGTTCGTCCT	CGATGGCCCTC	CACGTGAGGC	GTGTGGGAGG
60	41041	CGTAGTCGAC	CGCGATACGA	CGCACCCGCA	CCCCATCAGC	CTCATACCGC	GCCACCCACCT
	41101	CCTCCACCGC	CGACGGGTCC	CCCGCCACCA	CCGTCGAAGC	CGGACCATTA	CGCGCGCGA
	41161	TCCACACACC	CTCGACCAAGA	CCCACCTCAC	CGGCGGGCAA	CGCCACCGAA	GCCATGCC
	41221	CCCGGGCGGC	CAGCCGCGCC	GCGATCACCC	GAATGCGCAA	CGCCACCCACG	CGGGCGGC
	41281	CCTCCAGGCT	GAGGGCTCCG	GCCACACACG	CGGCCGGAT	CTCCCCCTGC	GAGTGTCCGA
	41341	CCACAGCGTC	CGGCACGACC	CCATGCGCCT	GCCACAGCGC	GGCCAGGCTC	ACCGCGACCG

41401 CCCAGCTGGC CGGCTGGAC ACCTCCACCC GCTCCGCCAC ATCCGACCGC GACAACATCT
 41461 CCCGCACATC CCAGCCCCTG TGCGGCAACA ACGCCCGCGC ACACTCCTCC ATACGAGCCG
 41521 CGAACACCAC GGAACCGGTCC ATGAGTTCCA CGCCCCATGCC CACCCACTGG GCACCCCTGCC
 41581 CGGGGAAGAC GAACACCGTA CGCGGCTGAT CCACCGCCAC ACCCATCAC CGGGCATCAC
 5 41641 CCAGCAGCAC CGCACCGGTGA CGAAGACAG CACGCTCACG CACCAACCCC TGCGCGACCG
 41701 CGGCCACATC CACCCCAACCC CGCGCAGAT ACCCCCTCCAG CCGCTCCACC TGCCCCCGCA
 41761 GACTCACCTC ACCACGAGCC GACACGGCA ACGGCACCAA CCCATCACCA CCCGACTCCA
 41821 CACGCGACGG CCCAGGAACA CCCTCAGGA TCACGTGCGC GTTCGTACCG CTCACCCCGA
 41881 ACGACGACAC ACCCGCATGC GGTGCCCGAT CCGACTCGGG CCACGGCCTC GCCTCGGTGA
 10 41941 GCAGCTCCAC CGCACCGGCC GACCAGTCCA CATGCGACGA CGGCTCGTCC ACGTGCAGCG
 42001 TCTTCGGCGC GATCCCATGC CGCATGCCA TGACCATCTT GATGACACCG GCGACACCCG
 42061 CAGCCGCCCTG CGCATGACCG ATGTTGACT TGACCGAAC GAGGTAGAGC GGCGTGTGCG
 42121 GGTCTGCCC GTAGGCCGCG AGGACGGCCT GCGCCTCGAT CGGGTCGCC AGCCGCGTGC
 42181 CGGTGCCGTG CGCCTCCACC ACGTCCACAT CGCGGCCGCG CAGTCCGGCG TTGACCAACG
 15 42241 CCTGCCGGAT CACCGCCTGC TGGGCGACGC CGTTGGGGC GGACAGTCCG TTGGAGGCAC
 42301 CGTCCTGGTT CACCGCCGAG CGCGGACGA CGCGAGAAC GGTGTGCCCG TTGCGCTCGG
 42361 CGTCGGAGAG CGCCTCCAGC ACGAGAACGC CGACGCCCTC GCGAAGACCG GTCCCGTCCG
 42421 CGCGTCGGC GAACGCCATTG CACCGTCCGT CGGGGGAGAG TCCGCGCTGC CGGGAGAACT
 42481 CCACGAGCTC TGCGGTGTTG GCCATGACGG TGACACCGCC GACCAGCGCC AGGGAGCACT
 20 42541 CCCCGGCCCG CAGTGCCTGT GCGCCTGGT GCAGGGCGAC CAGCGACGAC GAGCACGCCG
 42601 TGTCGACCGT GACCGCCGGG CCCTGAAGTC CGTACACGTA CGAGAGGCCG CGGACAGGA
 42661 CGCTCGTCTG CGTCGCCGTG ACACCGAGCC CGCCCAAGGTC CGGGCCGACG CGTAGCCCT
 42721 GTTGACACGC GCCCATGAAC ACGCCGGTGT CGCTCTCCCG GAGCTGTCC GGCACGATGC
 42781 CGGCCTCTC GAACGCCCTC CAGGAGGTCT CCAGGATCAG GCGCTGCTGG GGGTCCATCG
 25 42841 CCAGCGCTC GTTCGGACTG ATGCCAAGA ACGCGCGTC GAACCCGGCG CGGGCCAGGA
 42901 ATCCCGCGTG GCGTGTGCG GAGCGGCCGG CGCGTCCCGG GTCCGGTCG TACAGCGCGT
 42961 CGACGTCCC GCCCCGGTCG GTGGGAACT CGGTGATCGC CTCGGTACCG GCGCGACGA
 43021 GCGGCCACAG GTCCCTCCGC GAGGGCACCC CGCCGGGCAG TCGGCACGCC ATGCCGACGA
 43081 TCGCGACGGG GTCGCCGGAG CGAGGGGTCT GGGCGGTGCG GGGTGCCTG GTCGCGGAGC
 30 43141 CGCGGAGGTG GGCAGCGAAC GCACCGGGAG TGGGGTGGTC GAACCGGGTT GACCGGGCA
 43201 CCCGCAGACC CGTCCCGCGC GCGACGGTGT TGGTGAACCTC GACGGTGGT AGCGAGTCGA
 43261 GGCGTTCTC GCGGAACGTT CGGTCCGGGG AGCAGTGTCC GGCGCCCGGC AGGCCCCAGGA
 43321 CGGTGGCGAC GCTGTGCGG ACCAGTCGA CAGTACGTC CTCCCGGCC GCACGGGGCG
 43381 CGCGGAGGCC GTTCGCCAAC TCCTGTTCCG TGGCGTCGGG CTCGGCCGGT CGGTCAGTG
 35 43441 CGGTGAGGAT CGGGCGCGTG CGGCCCGCCA TCGTCCGGC CGGGCCCCCG CGGGAACCCG
 43501 TCCGGGCCAC GATGTACGAG CGCGCCCGCC CGATGGCCTT CTCGATCAGG TCGCCGGTGA
 43561 GCGCCGCCCG TTCGATGCCG GGCAGCGCGC GGACGGTGAC GGTGGGGAGT CCCTCCGCG
 43621 CCCGTGGCGC GGTGTGGCG TCGGCGCCGG CGGGGCCGTC GAGCAGGACG TGACAGAGCG
 43681 CGCCGGGGTT CGGGCTTCC TC GGCGTGC GGTCACGTG GGTGAGGCCG GTCTCGTCCG
 40 43741 GGAGCAGGCC GGCAGCGGTG TC GGCGTCC CCGGGTGAC CAGGACGGC GCGTCCGGC
 43801 CGATCGGAGG CGGCACGGTG AGGACCATCT TGCCGGTGTG CGGGCGTGG CTCATCCACG
 43861 CGAACCGTC CGCGCACGG CGGATGTCCC ACGGCTGAC CGGCAGCGGG CACAGCTCAC
 43921 CGCGTCGAA CAGGTCGAGG AGCAGTCGA GGATCTCCCG CAGGCGCGCG GGATCCACGT
 43981 CGGCCAGGTC GAACGGCTGC TGGGGCGGT GGCGGATGTC GGTCTTGC CCGGACCGA
 45 44041 ACCGCCGCC CGGTGCGAGC AGGCCGATGG ACGCGTCGAG GAGTCACCG GTGAGCGAGT
 44101 TGAGCACGAC GTCGACCGGC GGGAAAGGTGT CGCGAAGCC CGCGCTGCGG GAGTTGCCA
 44161 CATGGTCGGT GTCGAAGCCG TC GGCGTGC CGAGGTGTG TTTGGCGGG ACGAGGCCGT
 44221 CGTACACCTC GGCAGCGAGG TGGCGGGCGA TCCGGGTCGC CGCCATGCCG ACACGCCCG
 44281 TCGCGCGTG GACCAGGACC TTCTGGCCG GTCGCAGCTC GCGCGCTCG ACGAGGCCGT
 50 44341 ACCAGGGCGT GCGGAACACG ATGGGCACGG ACGCGCGAT GGGGAACGAC CATCCCCGTG
 44401 GGATCCGTGC GACCAGGCC CGGTCCCGA CA CGCTGCG CGGGAAACGCG TCCTGCACGA
 44461 GACCGAACAC GCGGTGCGCC GGGGCCAGGT CGTCGACGCC GGGTCCGACT TCGGTACGA
 44521 TGCCCGCCGC CTCCCCGCC ATCTCGCCCT CGCCCGGGTA GGTGCCGAGC GCGATCAGCA
 44581 CGTCGCGGAA GTTCAGCCCC GCGGCCGGA CGTCGATGCG GACCTCGCCG CGGGCCAGGG
 55 44641 CGCGGGCGGG ACGTGAGCG GGGCGACGAC GAGGTGCGGG AGCCTTCCGG AGGCGGGCG
 44701 CGCGACGCC CACTGGCGCG GTCGGCAGGG GGGTGGTGT CGCGCGTAC AGCCGGJGCA
 44761 CGTAGGCCAC CGCGGCCCGC AGCGCGATCT GGGGTTCGCC GAGCGAGGCC CGGGCGGGGA
 44821 CGAGGTCGTC ATCGCCGTCC GTGTCCACCA GCACGAACGA TCCGGGTCG CGGGCCTGGC
 44881 CGCGCAGCGC CTCGTCCCAG AGCCGGCCT GGTCCGCGTC CGGGATCTCG CGCGGGCCGA
 60 44941 CGCCCACCGC GCGGCGGGTG ACGACCGTCC GGCAGGGTGA CGGGGTGCCG GGCAGGTGCG
 45001 GCGCTCCCA GACCAGTTCG CACAGCGTGG CCTCGCCACT GCGCGTGGCG ACCAGATGGG
 45061 CGGGCAGCCC CGCGAGCCGC GCGCGCTGGA CTTGCCCCGA CGCGCGTGC GGATCGTGG
 45121 TGACGTGCCA GATCTCGTGC GGCACCTTGA AGTAGGCGAG CGCGCGGCCG CACTCGGCGA
 45181 GGATCGCTC GGCAGGGACG CGGGGGCCGT CGGAAACGAC GTAGAGCACG GGTATGTCG

45241 CGAGGACGGG GTGCCGGCGG CCCGCCGCGG CGGCCTCCCG GACACCGGCC ACCTCCTGGG
 45301 CGACGGTCTC GATCTCCCGG GGGTGGATGT TCTCCCCGCC GCGGATGATC AGCTCCTTGA
 45361 CCCGGCCGGT GATCGTCACG TGTCCGGTCT CGGCCTGACG TGCGAGGTCC CCGGTGCGGT
 45421 ACCAGCGTC CACGAGCACC TGGCGGTGCG CCTCCGGCTG GGCGTGGTAG CCGAGCATGA
 5 45481 GGCTCGGCC GCTCGCCAC AGCTCGCCCT CCTCGCCGGG TGCCACGTCG GCGCCGGACA
 45541 CGGGTGCAC GAACCGCAGC GACAGGCCCG GCACGGGCAG CCCGCACGAG CCGGAAACCC
 45601 GCGCATCCTC CAGGGTGTG GCGGTGAGCG AGCCGGTCGT CTCGGTGCAG CCGTACGTGT
 45661 CGAGCAGGGG CACGCCGAAC GTCGCCTCGA AATCCCTGGT GAGCGACGCC GGCGAGGTGG
 45721 ATCCGGCAGC CAGCGCCACG CGCAGCGCAG GAGCCCGCGG CTCGCCGGAC ACGGCGCCGA
 10 45781 GGAGGTAGCG GTACATCGTC GGCACGCCGA CGAGCACGGT GCTGGAGTGT TCGGCCAGGG
 45841 CGTCGAGGAC GTCACGCCG ACGAAGCCGC CCAGGATACG GGCGGACGCG CCGACCGTGA
 45901 GGACGGCGAG CAGGCAGAGG TGGTGGCCGA GGCTGTGGAA CAGCGGGGCG GGCCAGAGCA
 45961 GTTCGTGTC CTCGGTCAGC CGCCAGGACG GCACGTCGCA GTGCATCGCG GACCACAGGC
 15 46021 CGCTGCCTG TGCGGAAACC ACGCCCTTG GACGGCCGGT GGTGCCGGAG GTGTAGAGCA
 46081 TCCAGGCCGG TTCGTCCAGG CCGAGGTCTG CGCGGGCGG GCACGCCGGC TCGGTCCCAGG
 46141 CGAGGTCTC GTAGGAGACG CAGTCCGGT CCCGGCGCC GACGAGCAGC ACGGTGGCGT
 46201 CGGTGCCGGT GCGCGCACC TGGTCGAGGT GGGTTTCGTC GGTGACCAGC ACGGTCCGCG
 46261 CGGAGTCCGT CAGGAAGTGG GCGAGTCGG CGTCGGCGC GTCCGGGTTG AGCGGGACCG
 20 46321 CGACGGCGGC GGCGCGGGCG GCGCGGAGGT AGACCTCGAT GGTCTCGATC CGGTTGCCGA
 46381 GCAGCATCGC GACCCGGTCG CGCGGGTCGA CGCCGGACGC GGCGAGGTGT CCGGCGAGCC
 46441 GGCGGGCCCG GAGCCGGAGT TGCCTGTAACG TCACGGCCCG TTGGGAATCC GTGTAGGCAG
 46501 TCCGGTCGCC CGCTCGCTCG GCATGGATGC GGAGCAATTG GTGCAACGGC CGGATTGGTT
 46561 CCACACGCGC CATGGAAACA CCTTCTCTC GACCAACCGC ACAACAGCAC GGAACCGGCC
 46621 ACAGTAGAC GCCGGCGACG CTAGCAGCGT TTCCGGGACC GCCACCCCT GAAGATCCCC
 25 46681 CTACCGTGGC CGGCCCTCCCC GGACGCTCAT CTAGGGGTT GCACGCATAC CGCGTGCCTG
 46741 ATTGCCCTC CTGATGACCG ATGCCGGACG CCAGGGAGG GTGGAGGGCGT TGTCCATATC
 46801 TGTCAAGCGC CGGTATTGCC GCTTCGAGAA GACCGGATCA CGGACCTCG AGGGTGACGA
 46861 GACGGTGCTC GGCCTGATCG AGCACGGAC CGGCCACACC GACGTGTCGC TGGTGGACGG
 46921 TGCTCCCCGG ACCGCCGTGC ACACCAACGAC CGTGACCGAC GAGGCCTCA CCGAGGTCTG
 30 46981 GCACGCACAG CGCCCTGTGAGTCCGGCAT GGACAAACGGC ATGCCCTGGG CCGCACCGA
 47041 CGCGTACCTG TTCGGTGTG TGCGCACCGG CGAGAGCGGC AGGTACGCCG ATGCCACCGC
 47101 GGCCCTCTAC ACGAACGTC TCCAGCTCAC CGCGTGCCTG GGGTATCCCC TGCTCGCCCG
 47161 GACCTGGAAC TACGTCAAGCG GTATCAACAC GACGAACCGC GACGGGCTGG AGGTGTACCG
 47221 GGACTTCTGC GTGGGCCGGC CCCAGGCGCT CGACGAGGGC GGGATCGACC CGGCCACCAT
 35 47281 GCCCGGGGCC ACCGGTATCG GCGCCCACGG GGGCGGCATC ACCTCGTGT TCCTCGCCGC
 47341 CGGGGGCGGA GTCGGGATCA ACATCGAGAA CCCCAGCGTC CTACGCCCG ACCACTACCC
 47401 GACGACGTAC GGTCCGGGC CCCCCGTCTT CGCACGGGCC ACCTGGCTGG GCGCGCCGCA
 47461 GGGGGGCCGG CTGTTCATCT CGCGCACGGC CGGCATCCCT GGACACCGAA CGGTGCACCA
 47521 CGGTGATGTG ACCGGCCAGT GCGAGGTGCG CCTCGACAAAC ATGCCCGGG TCATCGCGC
 40 47581 GGAGAACCTG CGCGGCCACG CGTCCAGCG GGGGCACGTC CTGCCGACG TGGACCACT
 47641 CAAGGTCTAC GTCCGGCCGGC GCGAGGATCT CGATACGGTC CGCCGGTCT GCGCGCACG
 47701 CCTGTCGAGC ACCCGGGCCG TCGCCCTTTT GCACACCGAC ATAGCCCGCG AGGATCTGCT
 47761 CGTCGAAATC GAAGGCATGG TGGCGTACA ATACCGGTA AAAGGCCCGC GACGCTGCAGC
 47821 CTCGGGGAT CGCGAAGAG AAAGAAGAGC GTCACCGCAC AGCGCGGCAG CCGGTCCCT
 45 47881 TCGTCCTTCG CACAGCGGCC GATCTGGTTT CTCCAGCAAT TGGACCCGGA GAGCAACGCC
 47941 TATAATCTCC CGCTCGTGCACGCGCTCGC GGTCTATTGG ACACGCCGGC CCTGGAGCGT
 48001 GCGCTGGCGC TCGTCGTGCG GCGCACCGAG GCGTTGCCGA CGGTGTTCGA CACCGCCGAC
 48061 GCGGAGCCCG TCCAGGGGT GCTTCCCGCC CGGAAACACC TCCCTGCGCCA CGCGCGGGCG
 48121 GGCAGCGAGG AGGACGCCGC CGGGCTCGTC CGCGACGAGA TCGCCGCGCC GTTCGACCTC
 50 48181 GCCACCGGGC CGTTGATCAG GGCCCTGTG ATCCGCCCTCG GTGACGACGA CCACGTTCTC
 48241 GCGGTGACCG TGCACCATGT CGCCGGCGAC GGCTGGTCGT CGGGCTCTCT CCAACATGAA
 48301 CTCGCAGCCC ACTACACGGC GCTGCGCACG ACTGCCGCC CTGCCGAACCT GCGCCGTTG
 48361 CCGGTGCACT ACGCGCACTT CGCCGCCCTGG GAGCGGGCGC AACTCACCGG CGCCGGACTG
 48421 GACAGCGTC TGGCTACTG GCGCGAGCAA CTCCGGGGCG CCCCCGGCGC GCTCGCCCTC
 55 48481 CCCACCGACC GTCCCCGGCC CGCGTGCAGC GACCGGGACG CGGGCATGGC CGAGTGGCGG
 48541 CGCGGGCCCG CGCTGGCCAC CGCGTCCTC ACGCTCGCGC GCGACTCCGG TCGCTCCGTG
 48601 TTCAATGCC TCGTGGCGC CTTCCAAGCG GTCTCGGCC CGCAGGCCGG CACCGCGGAC
 48661 GTGCTGGTGC GACGCCCGT GGCGAACCGT ACACGCCGG CGTACGAGGG CCTGATCGGC
 48721 ATGTTCTGCA ACACGCTCGC GCTGCCGGC GACCTCTCGG CGATCCGTC GTTCCGGAA
 60 48781 CTCCCTCGAC GCTGCCGGGC CACGACCGAC GACCGCTTCG CCCACGCCGA CCTGCCGTT
 48841 GAGAACGTCA TCGAACCTCGT CGCACCGGAA CGCGACCTGT CGGTCAACCC GGTGTCCTCAG
 48901 GTGCTGGTGC AGGTGCTGCG CGCGACGCCG GCGACGCCGG CGCTGCCGG CATCGGGCC
 48961 GAACCGTTCC CGACCGGACG CTGGTTCAC CGCTCGACC TCGAATTCCA TGTGTACGAG
 49021 GAGCGGGGTG CGCGCGCTGAC CGCGCAACTG CTCTACAGCC GTGCGCTGTT CGACGAGCCA

49081	CGGATCACGG	GGTTGCTGGA	GGAGTTCACG	GCGGTGCTTC	AGGCGGTCAC	CGCCGACCCG	
49141	GACGTACGGC	TGTCGCGGCT	GCCGGCCGGC	GACCGCACCG	CGGCAGCGCC	CGTGGTCCCC	
49201	TCGAACGACA	CGGCGCGGGA	CCTGCCCGTC	GACACGCTGC	CGGGCCTGCT	GGCCCGGTAC	
49261	GCCGCAACGCA	CCCCCGGGCGC	CGTGGCCGTC	ACCGACCCGC	ACATCTCCCT	CACCTACGCG	
5	49321	CAGCTGGACC	GGCGGGCGAA	CCGCCTCGCG	CACCTGCTCC	GCGCGCGCGG	CACCGCCACC
	49381	GGCGACCTGG	TGGGGATCTG	CGCCGATCGC	GGCGCCGACC	TGATCGTCGG	CATCGTGGGG
	49441	ATCCTCAAGG	CGGGCGCCGC	TTATGTGCCG	CTGGACCCCG	AACATCCTCC	GGAGCGCACG
	49501	CGTTCGTGC	TGGCCGACGC	GCAGCTGACC	ACGGTGGTGG	CGCACGAGGT	CTACCGTTCC
	49561	CGGTTCCCCG	ATGTGCCGCA	CGTGGTGGCG	TTGGACGACC	CGGAGCTGGA	CGGGCAGCCG
10	49621	GACGACACGG	CGCCGGACGT	CGAGCTGGAC	CGGGACAGCC	TCGCCTACGC	GATCTACACG
	49681	TCCGGGTCGA	CCGGCAGGCC	GAAGGCCGTG	CTCATGCCGG	GTGTCAGCGC	CGTCAACCTG
	49741	CTGCTCTGGC	AGGAGCGCAC	GATGGGCCGC	GAGCCGGCCA	GCCGCACCGT	CCAGTTCTGT
	49801	ACGCCCACGT	TCGACTACTC	GGTGCAGGAG	ATCTTTCCG	CGCTGCTGGG	CGGCACGCTC
	49861	GTCATCCCGC	CGGACGAGGT	GCCTTGCAC	CCGCCGGGAC	TCGCCCGGTG	GATGGACGAA
15	49921	CAGGCATTAA	CCCGGATCTA	CGCGCCGACG	GCCGTACTGC	GCAGCCTGAT	CGAGCACGTC
	49981	GATCCGCACA	GCGACCAGCT	CGCCGCCCTG	CGGCACCTGT	GCCAGGGCGG	CGAGGCCTG
	50041	ATCCTCGACG	CGCGGTTGCG	CGAGCTGTGC	CGGCACCGGC	CCCACCTGCG	CGTGCACAAT
	50101	CACTACGGTC	CGGCCGAAAG	CCAGCTCATC	ACCGGGTACA	CGCTGCCCGC	CGACCCCGAC
	50161	CGCTGGCCCG	CCACCGCACC	GATCGGCCC	CCGATCGACA	ACACCCGCAT	CCATCTGCTC
20	50221	GACGAGGCAGA	GTGCGGCCGGT	TCCGGACGGT	ATGCCGGGGC	AGCTCTGCGT	CGCCGGCGTC
	50281	GGCCTCGCCC	GTGGGTACCT	GGCCCGTCCC	GAGCTGACCG	CCGAGCGCTG	GGTGCCGGGA
	50341	GATGCGGTG	GCGAGGAGCG	CATGTACCTC	ACCGGGCGACC	TGGCCCGCCG	CGCGCCCGAC
	50401	GGCGACCTGG	AATTCTCGG	CCGGATCGAC	GACCAAGTCA	AGATCCGCGG	CATCCGCGTC
	50461	GAACCGGGTG	AGATCGAGAG	CCTGCTCGCC	GAGGACGCC	CGCTCACGCA	GGCGCGGTG
25	50521	TCCGTGCGCG	AGGACCGGGC	GGGCGAGAAC	TTCCTGGCCG	CGTACGTCGT	ACCGGTGGCC
	50581	GGCCGGCACG	GCGACGACTT	CGCCGCGTCG	CTGCGCGCGG	GACTGGCCGC	CCGGCTGCC
	50641	GCCGCGCTCG	TGCCCTCCGC	CGTCGTCCGT	GTGGAGCGAC	TGCCGAGGAC	CACGAGCGGC
	50701	AAGGTGGACC	GGCGCGCGCT	GCCCCACCCG	GAGCCGGGCC	CGGCCTCGAC	CGGGGCGGTT
	50761	ACGCCCCGCA	CCGATGCCGA	CGGGACGGTG	TGCGGATCT	TCCAGGAGGT	GCTCGACGTC
30	50821	CCGCGGGTCG	GTGCCGACGA	CGACTCTTC	ACGCTGGCG	GGCACCTCC	GCTCGCCACC
	50881	CGGGTCTCT	CCGCATCCG	CGCCGAGCTG	GTCGGCGATG	TCCCCTCGC	TACGCTCTTC
	50941	GACGGGCGGA	CGCCCGCCGC	GCTCGCCCGT	CGGGCGGACG	AGGCCGGCCC	GGCCGCCCTG
	51001	CCCCCGATCG	CGCCCTCCGC	GGAGAACCGG	CGGGCCCCCC	TCACCGCGGC	ACAGGAACAG
	51061	ATGCTGCACT	CGCACGGCTC	GCTGCTCGCC	GGCCCGCTC	ACACGGTCGC	CCCGTACGGG
35	51121	TTCCGGCTGC	GGGGGCCACT	CGACCCGAA	GGCCTCGACG	CGGCACTGAC	CCGGATCGCC
	51181	GCGGCCACG	AGCCGCTGCC	GACCGGGTTC	CGCGATCGGG	AACAGGTCGT	CCGGCCGCC
	51241	GCTCCGGTGC	GCGCCGAGGT	GGTTCGGTG	CGGTCGGCG	ACGTCGACGC	CGCGGTCCGG
	51301	GTCGCCAAC	GGGAGCTGAC	CCGGCCGTTG	GACCTCGTGA	ACGGGTCGTT	GCTCGTGC
	51361	GTGCTGCTGC	CGCTGGCGC	CGAGGATCAC	GTGCTGCTGC	TGATGCTGCA	CCACCTCGCC
40	51421	GGTGACGGAT	GGTCCTTCGA	CCTCTGGTC	CGGGAGTTG	CGGGGACGCA	ACCGGACCTT
	51481	CCGGTGTCT	ACACGGACGT	GGCCCCGTGG	GAACGGAGTC	CGGCCGTGAT	CGCGGCCAGG
	51541	GAGAACGACC	GGGCCTACTG	GCGCCGGCGG	CTGGGGGGCG	CCACCGCGCC	GGAGCTGCC
	51601	GCGGTCCGGC	CCGGCGGGGG	ACCGACCGGG	CGGGCGTTCC	TGTGGACGCT	CAAGGACACC
	51661	GCCGTCTGG	GGGCACGCCG	GGTCGCGGAC	GCCCACGACG	CGACGTTGCA	CGAAAACCGTG
45	51721	CTCGGCGCT	TCGCCCTGGT	CGTGGCGGAG	ACCGCCGACA	CCGACCGACGT	GCTCGTCGCG
	51781	ACGCCGTTCG	CGGACCGGGG	GTACGCCGGG	ACCGACCAACC	TCATCGGCTT	CTTCGCGAAG
	51841	GTCCTCGCGC	TGCGCCTCGA	CCTCGCGGCC	ACGCCGTCGT	TCCCCGAGGT	GCTCGCCGG
	51901	CTGCACACCG	CGATGGTGGG	CGCGCACGCC	CACCAGGCGG	TGCCCTACTC	CGCGCTGCC
	51961	GCCGAGGACC	CCCGCGCTGCC	GCCGGCCCCC	GTGTCGTTCC	AGCTCATCAG	CGCGCTCAGC
50	52021	GCGGAACCTGC	GGCTGCCCGG	CATGCACACC	GAGCCGTTCC	CCGTCGTCGC	CGAGACCGTC
	52081	GACGAGATGA	CCGGCGAAC	GTCGATCAAC	CTCTCGACG	ACGGTCGCAC	CGTCTCCGGC
	52141	GCGGTGGTCC	ACGATGCCGC	GCTGCTCGAC	CGTGCCACCG	TCGACGATT	GCTCACCCGG
	52201	GTGGAGGCGA	CGCTGCGTGC	CGCCGCGGGG	GACCTCACCG	TACGCGTCAC	CGGTTACGTG
	52261	GAAAGCGAGT	AGCCATGCC	GAGCAGGACA	AGACAGTCGA	GTACCTTCGC	TGGGCGACCG
55	52321	CGGAACCTCCA	GAAGACCCGT	CGGAAACTCG	CCGCGCACAG	CGAGCCGTTG	GCGATCGTGG
	52381	GGATGGCCTG	CCGGCTGCC	GGCGGGGTGCG	CGTCGCCGGA	GGACCTGTGG	CAGTTGCTGG
	52441	AGTCCGGTGG	CGACGGCATC	ACCGCGTTCC	CCACGGACCG	GGGCTGGGAG	ACCACCGCCG
	52501	ACGGTCGCGG	CGGCTTCCTC	ACCGGGGCGG	CCGGCTTCGA	CGCGCGTTC	TTCGGCATCA
	52561	GCCCAGCGCA	GGCGCTGGCG	ATGGACCCGC	AGCAGCGCCT	GGCCCTGGAG	ACCTCGTGGG
60	52621	AGGCAGTTCGA	GCACGCGGGC	ATCGATCCGC	AGACGCTGCG	GGGCAGTGAC	ACGGGGGTGT
	52681	TCCTCGCGC	GTTCTCCAG	GGGTACGGCA	TCGGCGCCGA	CTTCGACGGT	TACGGCACCA
	52741	CGAGCATTCA	CACGAGCGTG	CTCTCCGGCC	GCCTCGCGTA	CTTCTACGGT	CTGGAGGGTC
	52801	CGGCGGTCAC	GGTCGACACG	GGGTGTTCGT	CGTCGCTGGT	GGCGCTGCAC	CAGGCCGGGC
	52861	AGTCGCTGCG	CTCCGGCGAA	TGCTCGCTCG	CCCTGGTCGG	CGGCCTACG	GTGATGCC

52921 CGCCGGCGGG GTTCGCGGAC TTCTCCGAGC AGGGCGGCCT GGCCCCCGAC GCGCGCTGCA
 52981 AGGCCTTCGC GGAAGCGGCT GACGGCACCG GTTTCGCCGA GGGGTCCGGC GTCCCTGATCG
 53041 TCGAGAAGCT CTCCGACGCC GAGCGAACG GCCACCGCGT GCTGGCGGT GTCGGGGGTT
 53101 CCGCCGTCAA CCAGGACGGT GCCTCCAACG GGCTGTCCGC GCCGAACGGG CCGTCGCAGG
 5 53161 AGCGGGTGAT CCGGCAGGCC CTGGCCAACG CGGACTCAC CCCGGCGGAC GTGGACGCCG
 53221 TCGAGGCCA CGGCACCAGC ACCAGGCTGG GCGACCCCAT CGAGGCACAG GCGTGTCTGG
 53281 CCACCTACGG GCAGGGCGC GACACCCCTG TGCTGTCTGG CTCGCTGAAG TCCAACATCG
 53341 GCCACACCCA GGCGCCGCC GGCCTGCCCG GTGTCATCAA GATGGTCCTC GCCATGCCG
 53401 ACGGCACCCCT GCCCCGCACC CTGCACGTGG ACACGCCGTC CTCGCACGTC GACTGGACGG
 10 53461 CCGGCGCCGT CGAACTCCTC ACCGACGCCG GGCCTGGCC CGAAACCGAC CGCCCACGGC
 53521 GCGCCGGTGT CTCCCTCTTC GGCGTCAGCG GCACCAACGC CCACATCATC CTCGAAAGCC
 53581 ACCCCCCGACC GGCCCCCGAA CCCGCCCGG CACCCGACAC CGGACCGCTG CCGCTGTCTG
 53641 TCTCGGCCCG CACCCCGCAG GCACTCGACG CACAGGTACA CGCCTGCCG GCGTTCTCG
 53701 ACGACAACCC CGCGCGGGAC CGGGTCGCCG TCGCGCAGAC ACTCGCCCGG CGCACCCAGT
 15 53761 TCGAGCACCG CCGCGTGTG CTCGGCGACA CGCTCATCAC CGTAGGCCG AACGCCGGCC
 53821 GCGGACCGGT GTCTTCGTC TACTCGGGC AAAGCACCGT GCACCCGAC ACCGGGGCGC
 53881 AACTCGCTC CACCTACCCC GTGTCGCCG AAGCGTGGCG CGAGGCCCTC GACCACCTCG
 53941 ACCCCCCACCA GGCCCCGGCC ACGCACTTC CCCACAGAC CGCGCTCACC GCGCTCTGC
 54001 GGTCTGGGG CATCACCCCG CACGCCGTCA TCGGCCACTC CCTCGGTGAG ATCACCGCCG
 20 54061 CGCACGCCGC CGGTGTCTG TCCCTGAGGG ACGCCGGCGC GCTCTCACC ACCCGCACCC
 54121 GCCTGATGGA CCAACTGCCG TCGGGCGGCC CGATGGTCAC CGTCTGACC AGCGAGGAAA
 54181 AGGCACGCCA GGTCTGCCG CGGGGACGAG GAAGCCGTAC TCGAAGCCGC CGGCAGCTC GGCATCCACC
 54241 TCGTGTGTC CGGGGACGAG GAAGCCGTAC TCGAAGCCGC CGGCAGCTC GGCATCCACC
 54301 ACCGCCTGCC GACCCGCCAC GCGGCCACT CCGAGCGCAT GCAGCCACTC GTCGCCCCC
 25 54361 TCCTCGACGT CGCCCGGACC CTGACGTACC ACCAGCCCCA CACGCCATC CCCGGCGACC
 54421 CCACCAACGC CGAATACTGG GCGCACCAAG TCCCGCACCAG AGTACGTTTC CAGGCGCACA
 54481 CCGAGCAGTA CCCGGCGCG ACGTTCTCG AGATCGGCC CAACCAGGAC CTCTCGCCG
 54541 TCGTCACGG CGTTGCCGCC CAGACCGTA CGCCCGACGA GGTGGGGCG CTGCACACCG
 54601 CGCTCGCGCA GCTCCACGTC CGCGCGCTCG CGATCGACTG GACGCTCGTC CTCGGCGGGG
 30 54661 ACCGCCGCC CGTCACGCTG CCCACGTATC CGTTCCAGCA CAAGGACTAC TGGCTCGGGC
 54721 CCACCTCCCG GGCGATGTG ACCGGCGCGG GGCAGGAGCA GGTGGCGCAC CCGCTGCTCG
 54781 GCGCCGCGGT CGCGCTGCCG GGCACGGCG GAGTCGTCT GACCGGCCGC CTGTCGCTGG
 54841 CCTCCCATCC GTGGCTCGGC GAGCACCGGG TCGACGGCAC CGTCTCCTG CCCGGCGCGG
 54901 CCTTCCTCGA ACTCGCGGC CGCGCCGGCG ACGAGGTCTGG CTGCGACCTG CTGCACGAAC
 35 54961 TCGTCATCGA GACGCCGCTC GTGCTGCCG CGACCGGGG TGTGGCGTC TCCGTCGAGA
 55021 TCGCCGAACC CGACGACACG GGGCGCGGG CGGTCACCGT CCACGCGCGG GCGACGGCT
 55081 CGGGCTGTG GACCCGACAC GCGGCCGGAT TCCCTGGCAC GGCACCCGGCA CGGGCACCG
 55141 CCACGGACCC GGCACCCCTGG CGGCCCGCGG AAGCCGGACC GGTGACGTC GCGACGCTC
 55201 ACGACCGGTT CGAGGACATC GGGTACTCCT ACGGACCGGG CTTCCGGGG CTGCGGCCG
 40 55261 CCTGGCGCGC CGCGACACC GTGTACGCCG AGGTCGCGCT CCCCGACGAG CAGAGCGCCG
 55321 ACGCCGCCCG TTTCACGCTG CACCCCGCGC TGCTCGACGC CGCGTCTCCAG CGCGCGCGC
 55381 TGGCCGCGCT CGACGACACC GGCAGGGCGG CCCGACTGCC GTTCTCGTTC CAGGACGTCC
 55441 GCATCCACGC GGCGGGGGCG ACGCGCTGC GGGTCACGGT CGGCCGCGAC GGCAGGGCA
 55501 GCACCGTCCG CATGACCGGG CGGGACCGGG AGCTGGTGGC CGTGGTCGGT CCGGTGCTGT
 45 55561 CGCGCCCGTA CGCGGAAGGC TCCGGTGACG CCCTGCTGCG CCCGGTCTGG ACCGAGCTGC
 55621 CGATGCCCGT CCCGTCGCCG GACGATCCGC CGCTGGAGGT CCTCGCGGCC GACCCGGCG
 55681 ACGGCGACGT TCCGGCGGCC ACCCGGGAGC TGACCGCCCG CGTCTCGCGC CGCTCCAGC
 55741 GCCACCTGTC CGCCGCCGAG GACACCACCT TGGTGGTACG GACCCGGCACC GCCCCGGCG
 55801 CTGGCGCCGC CGCGGGCTGTG GTCCGCTCGG CGCAGGCCGA GAACCCGGC CGCGTCGTGC
 50 55861 TCGTCGAGGC GTCCCCGGAC ACCTCGGTGG AGCTGCTCGC CGCGTGCGCC CGCGTCGTGC
 55921 AACCGCAGCT GGCGCTCCGG GACGGCGTGC TCTTCGCCG CGGGCTGGTC CGGATGTCGG
 55981 ACCCCCCGCA CGGCCCCGCT TCCCTGCCGG ACGGCGACTG GCTGCTCACC CGGTCCGCT
 56041 CGGGCACGTT GACGACGCTC GCGCTCATAG CGACGACAC GCCCCGGCGG GCGCTCGAAG
 56101 CGGGCGAGGT CGCGCATCGAC GTCCGCCGG CGGACTGAA CTTCCGCGAT GTGCTGATCG
 55 56161 CGCTCGGGAC GTACACCGGG GCCACGCCA TGGCGGGCA GGCGCGGGC GTCGTGGTGG
 56221 AGACCGGGCC CGGCGTGGAC GACCTGTCCC CGGGCGACCG GGTGTTCGGC CTGACCCGGG
 56281 GCGGCATCGG CCCGACGGCC GTCACCGACCG GGCGCTGGCT GGCGCGGATC CCCGACGGCT
 56341 GGAGCTTCAC CACGGCGGCC TCCGTCCGA TCGTGTTCGC GACCGCGTGG TACGGCCTGG
 56401 TCGACCTCGG CACACTGCCG GCCGGCGAGA AGGTCTCGT CCACGCGGCC ACCGGCGGTG
 60 56461 TCGGCATGGC CGCCGACACAG ATCGCCCGCC ACCTGGCGC CGAGCTCTAC GCCACCGCCA
 56521 GTACCGGCAA GACGACGCTC CTGCGCGCC CGGGGCTGCC CGACACGAC ATCGCCGACT
 56581 CTCGGACGAC CGCGTCTCGG ACCGCTTCC CGCGCATGGA CGTCGTCCTG AACGCGCTGA
 56641 CGGGCGAGTT CATGACCGCG TCGCTCGACC TGCTGGACGC CGACGGCGGG TTCGTCGAGA
 56701 TGGGCCGCAC CGAGCTGCCG GACCCGGCCG CGATCGTCCC CGCCTACCTG CGGTTCGACC

56761 TGCTGGACGC GGGCGCCGAC CGCATCGCG AGATCCTGGG CGAACTGCTC CGGCTGTTCG
 56821 ACGCGGGCGC GCTGGAGCCG CTGCCGGTCC GTGCCCTGGGA CGTCGGCGCAG GCACGGCAUCG
 56881 CGCTCGGCTG GATGAGCCG GCCCCCACCA TCGGCAAGAA CGTCCTGACG CTGCCCCGGC
 56941 CGCTCGACCC GGAGGGCGCC GTCGTCTCA CGGGCGGCTC CGGCACGCTC GCCGGCATCC
 5 57001 TCGCCCCGCCA CCTGCGCGAA CGGCATGTCT ACCTGCTGTC CGGAGCGGCA CGGCCCCGAGG
 57061 GGACGCCCCGG CGTCCACCTG CCCTGCGACG TCGGTGACCG GGACCAGCTG CGGGCGGCC
 57121 TGGAGCGGGT GGACCGGGCG ATCACCGCCG TGGTGCACCT CGCCGGTGC CGTGACGACG
 57181 GCACCGTCGC GTCGCTCACCC CGCGAGCGTT TCGACACGGT GCTGCGCCCG AAGGCGGACG
 57241 GCGCCTGGTA CCTGACAGAG CTGACGAAGG AGCAGGACCT CGCCCGTTC GTGCTCTACT
 10 57301 CGTCGGCCGC CGGCGTGCTC GGCAACGCCG GCCAGGGCAA CTACGTCGCC GCGAACCGGT
 57361 TCCTCGACGC GCTCGCCGAG CTGCCACAG GTTCCGGCT GCCGGCCCTC TCCATCGCCT
 57421 GGGGGCTCTG GGAGGACGTG AGCAGGGCTCA CGCGGGCGCT CGGCGAAGCC GACCGGGACC
 57481 GGATGCGGCG CAGCGGTTTC CGGGCCATCA CGCGCACA CGGCATGAC CGTACGAGG
 57541 CGGCCGGCCG CACCGGAAGT CCCGTGGTGG TCGCGGCGGC GCTGACGAC GCGCCGGACG
 15 57601 TGCCGCTGCT GCGCGGCCCTG CGGCGGACGA CGTCCGGCG GGCCGCCGTC CGGGAGTGT
 57661 CGTCCGCCGA CGCGCTCGCC GCGCTGACCG GCGACGAGCT CGCCGAAGCG CTGCTGACGC
 57721 TCGTCCGGGA GAGCACCGCC GCGTGCTCG GCCACGTGGG TGGCGAGGAC ATCCCCCGCA
 57781 CGGCGGCGTT CAAGGACCTC GGCATCGACT CGTCACCCGC GTTCCAGCTG CGCAACGCC
 57841 TCACCGAGGC GACCGGTGTG CGGCTGAACG CCACGGCGGT CTTGACTTC CCGACCCCGC
 20 57901 ACGTGCTCGC CGGGAAAGCTC GGCGACGAAC TGACCGGCAC CGCGCGGCC GTCGTGCC
 57961 GGACCGCGGC CACGGCCGGT GCGCACGACG AGCCGCTGGC GATCGTGGGA ATGGCCTGCC
 58021 GGCTGCCCGG CGGGGTGCGG TCACCCGAGG AGCTGTGGCA CCTCGTGGCA TCCGGCACCG
 58081 ACGCCATCAC GGAGTTCCCG ACGGACCGCG GCTGGGACGT CGACGCGATC TACGACCCGG
 58141 ACCCCCACGC GATCGGCAAG ACCTCGTCC GGACGGTGG CTTCTCACC GGCGCGACAG
 25 58201 GCTTCGACGC GGCGTTCTTC GGCATCAGCC CGCGCGAGGC CCTCGCGATG GACCCGCAGC
 58261 AGCGGGTGTG CCTGGAGACG TCGTGGGAGG CGTTCGAAAG CGCCGGCATT ACCCGGACT
 58321 CGACCCCGGG CAGCGACACC GGCGTGTGCG TCGCGCCCTT CTCTACGGT TACGGCACCG
 58381 GTGCGGACAC CGACGGCTTC GGCGCGACCG GCTCGCAGAC CAGTGTGCTC TCCGGCCGGC
 58441 TGTCGTACTT CTACGGTCTG GAGGGTCCGG CGGTACCGGT CGACACGGCG TGTTCGTGT
 30 58501 CGCTGGTGGC GCTGCACCA CGCGGGCAGT CGTCGCGTC CGCGAATGC TCGCTCGCCC
 58561 TGTCGGCGG CGTCACGGTG ATGGCGTCTC CGGGCGGCTT CGTGGAGTTC TCCCAGCAGC
 58621 GCGGCCCTCGC GCGGACGCG CGGGCGAAGG CGTTCGGCGC GGGTGGGAC GGCACGAGCT
 58681 TCGCCGAGGG TGCCGGTGTG CTGATCGTCG AGAGGCTCTC CGACGCCGAA CGAACCGGTC
 58741 ACACCGTCTT CGCGGTGCGT CGTGGTTCGG CGGTCAACCA GGATGGTGC TCCAACGGGC
 35 58801 TGTCGGCGCC GAACGGGCG CGCAGGAGG GGGTGATCCG CGACGCCCTG GCCAACGCC
 58861 GGCTCACCCCC CGCGGACGTG GACCCGCTCG AGGCCCACGG CACCGGCACC AGGCTGGGCG
 58921 ACCCCATCGA GGCACAGGGC GTACTGGCCA CCTACGGACA GGAGCGCGCC ACCCCCCCTGC
 58981 TGCTGGGCTC GCTGAAGTCC AACATCGGCC AGCCCGCAGGC CGCGTCCGGC GTCGCCGGCA
 59041 TCATCAAGAT GGTGCAGGGC CTCCGGCACG GGGAGCTGCC CGCCGACGCTG CACGCCGACG
 40 59101 AGCCGTCGCC GCACGTCAC TGGACGGCCG GCGCCGTCGA ACTGCTGACG TCGGCCGGC
 59161 CGTGGCCCGA GACCGACCGG CCACGGCGTC CGGCCGTCCTC CTGGTTGGG GTGAGCGGC
 59221 CCAACSCCCCA CGTCATCTCG GAGGGCGGAC CGGTAACCGGA GACCCCGCGC GCATCGCCTT
 59281 CCGGTGACCT TCCCTGCTG GTGTCGGCAC GCTCACCGGA AGCCGCTGAC GAGCAGATCC
 59341 GCGGACTGCG CGCCTACCTG GACACCA CGGACGTCGA CGGGGTGGCC GTGGCACAGA
 45 59401 CGCTGGCCCG GCGCACACAC TTGCCCCACC CGGCCGTGCT GCTCGGTGAC ACCGTCTATCA
 59461 CCACACCCCC CGCGGACCGG CCCGACGAAC TCGTCTCGT CTACTCCGGC CAGGGCACCC
 59521 AGCATCCCGC GATGGGCGAG CAGCTCGCC CGGCCCATCC CGTGTTCGCC GACGCCCTGGC
 59581 ATGAAGCGCT CGGCCGCCTT GACAACCCCG ACCCCCCACGA CCCACGCGAC AGCCAGCATG
 59641 TGCTCTTCGC CCACCAAGGGC GCGTTCACCG CCCTCTGCG GTCTCTGGGC ATCACCCGC
 50 59701 ACGCGGTCTA CGGCCACTCG CTGGCGAGA TCACCGCCG GCACGCCGCC GGCATCCTGT
 59761 CGCTGGACGA CGCGTGCACC CTGATCACCA CGCGCGCCCG CCTCATGAC ACGCTCCGC
 59821 CACCCGGTGC CATGGTCACC GTACTGACCA GCGAAGAGAA GGCAACGCCAG CGTGTGCGC
 59881 CGGGCGTGGA GATCGCCGCC GTCAACGGGC CCCACTCCAT CGTGTGTC GGGGACCGAGG
 59941 ACGCCGTGCT CACCGTCGCC GGGCAGCTCG GCATCCACCA CGGCCCTGCC GCCCCGCACG
 55 60001 CGGGGACTC CGCGCACATG GAGCCCGTGG CGGCCGAGCT GCTCGCCACC ACCCGCGGGC
 60061 TCCGCTACCA CCCTCCCCAC ACCTCCATTG CGAACGACCC CACCAAGCGT GAGTACTGGG
 60121 CCGAGCAGGT CCGCAAGCCC GTGCTGTTCC ACACCCACGC CGACGAGTAC CGGGACGCC
 60181 TGTCGGTGGA GATCGGCCCG GCGCAGGACC TCTCCCCGCT CGTCGACGGG ATCCCCGTG
 60241 AGAACGGCAC CGCGGACGAG GTGACCGCGC TGACACCCG GCTCGCGAC CTCTACCGC
 60 60301 GCGGTGCCAC GCTCGACTGG CCCCCCATCC TCGGGGCTGG GTCAACGGCAC GACGCGGATG
 60361 TGCCCGCGTA CGCGTTCCAA CGCGGGACT ACTGGATCGA GTCGGCACGC CGGGCCGCAT
 60421 CCGACCGGGG CCACCCCGTG CTGGGCTCCG GTATGCCCT CGCCGGGTGCG CGGGGCCGGG
 60481 TGTCACGGG TTCCGTGCCG ACCGGTGCAGG ACCGGCGGGT GTTCGTCGCC GAGCTGGCGC
 60541 TGGCCGCCGC GGACGCGGGTC GACTGCGCCA CGGTGAGCG GCTCGACATC GCCTCCGTGC

60601 CCGGCCGGCC GGGCCATGGC CGGACGACCG TACAGACCTG GGTCGACGAG CCGGCAGC
 60661 ACGGCCGGCG CCGGTTCAAC GTGCACACCC GCACCGGEGA CGCCCCGTGG ACGCTGCACG
 60721 CCGAGGGGGT GCTGCGCCCC CATGGCACGG CCCTGCCCAGA TGCGGCCGAC GCGAGTGGC
 60781 CCCCACCGGG CGCGGTGCCG CGGGACGGGC TGCCGGGTGT GTGGCGCCGG GGGGACCAGG
 5 60841 TCTTCGCCGA GGCGAGGTG GACGGACCGG ACGGTTCTGT GGTGCACCCCC GACCTGCTCG
 60901 ACGCGGTCTT CTCCGCGGTG GGCACGGAA GCCGCCAGCC GGCCGGATGG CGCGACCTGA
 60961 CGGTGCACGC GTCGGACGCC ACCGTACTGC GCGCCTGCCT CACCCGGCGC ACCGACGGAG
 61021 CCATGGGATT CGCCGCCCTC GACGGGCCCG GCCTGCCGGT ACTCACCGCG GAGGCGGTGA
 61081 CGCTGCCGA GGTGGCGTCA CGTCCGGCT CCGAGGAGTC GGACGGCCTG CACCGGTTGG
 10 61141 AGTGGCTCGC GTGCGCCGAG CGGGTCTACG ACGGTGACCT GCCCCGAGGA CATGTCTGA
 61201 TCACCGCCGC CAACCCCGAC GACCCCGAGG ACATACCCAC CCGCGCCCAC ACCCGCGCCA
 61261 CCCCGCTCT GACCGCCCTG CAACACCACC TCACCACAC CGACCACACC CTCATCGTCC
 61321 ACACCACAC CGACCCCGCC GGCGCCACCG TCACCGGCCT CACCCGCACC GCCCAGAACG
 61381 AACACCCCCA CCGCATCCGC CTCATCGAAA CCGACCAACCC CCACACCCCC CTCCCCCTGG
 15 61441 CCCAACCTCGC CACCCCTCGAC CACCCCCACC TCCGCTCAC CCACCAACACC CTCCACCA
 61501 CCCACCTTCAC CCCCCCTCCAC ACCACCAACCC CACCCACCCAC CACCCCCCTC AACCCCCGAAC
 61561 ACGCCATCAT CATCACCGGC GGCTCCGGCA CCCTCGCCGG CATCCTCGCC CGCCACCTGA
 61621 ACCACCCCCA CACCTACCTC CTCTCCCGCA CCCCCACCCCC CGACGCCACC CCCGGCACCC
 20 61681 ACCTCCCTG CGACGTGGC GACCCCCACC AACTCGCCAC CACCCCTCAC CTCCACGCC
 61741 AACCCCTTCAC CGCCATCTTC CACACCGCG CCACCCCTCGA CGACGGCATE CTCCACGCC
 61801 TCACCCCCGA CGGCCCTCAC ACCGCTCTCC ACCCCAAAGC CAACGCCGCC TGGCACCTGC
 61861 ACCACCTTCAC CCAAAACCAA CCCCCTCACCC ACTTCGTCT CTACTCCAGC CCCGCCJCCG
 61921 TCCTCGGAG CGCCCGACAA GGAAACTACG CGGCCGCCAA CGCCCTCCCTC GACGCCCTCG
 25 61981 CCACCCACCG CCACACCCCTC GGCCAACCCCG CCACCTCCAT CGCCTGGGGC ATGTGGACAA
 62041 CCACCAGCAC CCTCACCGGA CAACTCGACG AGCGCGACCG GGACCGCATC CGCCGCGGCC
 62101 GTTCTCTCCC GATCACGGAC GACGAGGGCA TGCGCTCTA CGAGGCGGCC GTCGGCTCCG
 62161 GCGAGGACTT CGTCATGGGC GCGCGATGG ACCCGGCACA GCCGATGACC GGCTCCGTAC
 62221 CGCCCATCTT GAGCGGCCCTG CGCAGGAGCG CGCGGCGCGT CGCCCGTGCC GGGCAGACGT
 62281 TCGCCCAGCG GCTCGCCGAG CTGCCCAGCG CCGACCGCGG CGCGGCGCTG ACCACCCCTG
 30 62341 TCTCGGACGC CACGGCCGGC GTGCTCGGCC ACGCCGACCG CTCCGAGATC GCGCGACCA
 62401 CGACGTTCAA GGACCTCGGC ATCGACTCGC TCACCGCGAT CGAGCTGCGC AACCGGCTCG
 62461 CGGAGGCGAC CGGGCTGGCG CTGAGTGCCA CGCTGGTGT CGACCAACCCG ACACCTCGGG
 62521 TCCCTCCCGC CAAGCTCCGC ACCGATCTGT TCGGCACCGC CGTGGCCACG CCCGCGCGGA
 62581 CGGCACGGAC CCACCAACGAC GAGCCACTCG CGATCGTGG CATGGCGTGC CGACTGCCCG
 35 62641 GCGGGGTCGC CTCGCGGGAG GACCTGTGGC AGCTCGTGGC GTCCGGCACC GACGCGATCA
 62701 CCGAGTTCCC CACCGACCGC GGCTGGGACA TCGACCGGCT GTTCGACCCG GACCCGGACG
 62761 CCCCCGGCAA GACCTACGTC CGGCACGGCG GCTTCCTCGC CGAGGCGGCC GGCTTCGATG
 62821 CCGCGTTCTT CGGCATCAGC CGCGCGGAGG CACGGGCCAT GGACCCGCAG CAGCGCGTCA
 62881 TCCCTGAAAC CTCTGGGAG GCGTTCGAGA ACGCGGGCAT CGTGCCGGAC ACGCTGCGCG
 40 62941 GCAGCGACAC CGCGTGTTC ATGGGCGCT TCTCCCATGG GTACGGCGCC GGCCTCGAC
 63001 TGGGCGGGTT CGGCGCCACC GCCACGAGA ACAGCGTGT CTCCGGCCGG TTGTCGIACT
 63061 TCTTCGGCAT GGAGGGCCCG GCGTCACCG TCGACACCGC CTGCTCGTGC TCGCTGGTCG
 63121 CCCTGACCA GGCGGCACAG GCGCTCGGA CTGGAGAAATG CTCGCTGGCG CTCGCCGGCG
 45 63181 GTGTCAACGGT GATGCCACCC CGCTGGGCT ACGTCGAGTT CTGCGGCCAG CGGGGACTCG
 63241 CCCCCGACGG CGGTTGCCAG GCCTTCGCGG AAGGCGCCGA CGGCACGAGC TTCTCGGAGG
 63301 GCGCCGGCGT TCTTGTGCTG GAGGGCTCT CCGACGCCGA CGCAACCGGA CACACCGTCC
 63361 TCGCGGTCTG CCGCTCTCC GCGCTCAACC AGGACGGCGC CTCCACCGGC ATCTCCGCAC
 63421 CCAACGGCCC CTCCCAGCAG CGCGTCATCC GCCAGGCCCT CGACAAAGGCC GGGCTCGCCC
 50 63481 CGGCCGACGT GGACGTGGT GAGGCCACG GCACCGGAAC CCCGCTGGGC GACCCGATCG
 63541 AGGCACAGGC CATCATCGCG ACCTACGGCC AGGACCGCGA CACACCGCTC TACCTCGGTT
 63601 CGGTCAAGTC GAACATCGGA CACACCCAGA CCACCGCCGG TGTCGCCGGC GTCATCAAGA
 63661 TGGTCATGGC GATGCGCCAC GGCATCGCGC CGAAGACACT GCACGTGGAC GAGCGCTCGT
 63721 CGCATGTGGA CTGGACCGAG GGTGCGGTGG AACTGCTCAC CGAGGCGAGG CCGTGGCCCG
 63781 ACGCGGGACG CCCGCGCCCG CGGGCGTGT CGTCGCTCGG TATCAGCGGT ACGAACGCC
 55 63841 ACGTGATCCT TGAGGGTGT CCCGGGCCGT CGCGTGTGGA GCCGCTGTGTT GACGGGTITGG
 63901 TGCGGTTGCC GGTGCGGTG CGGAGTGGAG CGAGTCTGCC GGGGCCAGGTG GAGCGGCTGG
 63961 AGGGGTATCT GCGCGGGAGT GTGGATGTGG CGCGCGTCCG GCAGGGGTTG GTGCCTGAGC
 64021 GTGCTGTCTT CGGTACCGGT CGGGTACTGC TGGGTGATGC CGGGGTGATG GGTGTGGCG
 64081 TGGATCAGCC CGGTACGGTG TTCGTTTC CGGGCGAGGG TGCTCAGTGG GTGGGCATGG
 60 64141 GTGTGGAGTT GATGGACCGT TCTGCGGTG TCGCGCTCG TATGGAGGAG TGTGCGCGGG
 64201 CGTTGTTGCC GCACACGGGC TGGGATGTGC GGGAGATGTT GGCACGCCCG GATGTGGCG
 64261 AGCGGGTGGG GGTGGTCCAG CGGGCCAGCT GGGCGGTCCG GTCAAGCCTG GCCGCACGT
 64321 GGCAGGCCA CGGGGTGCGTA CCCGACCGGG TGATCGGACA CTCCCAAGGGC GAGATCGGG
 64381 CGGCCTGCGT GGCGGGGGCC CTCAGCCTG AGGACGCCGC CGCGGTGGTG GCCTTGCGCA

64441	GCCAGGTCAT	CGCGGCCGGA	CTGGCCGGGC	GGGGAGCGAT	GGCTTCGGTG	GCATTGCCGG	
64501	CCGGTGAGGT	CGGTCTGGTC	GAGGGCGTGT	GGATCGCCGC	GCGTAACGGC	CCCGCCTCGA	
64561	CAGTCGTGGC	CGCGGAGCCG	TCGGCGGTGG	AGGACGTGGT	GACCGGGTAT	GAGACCGAAG	
64621	GCGTCCGAGT	CGCTCGTATC	GCCGTCGACT	ACGCCTCCA	CACGCCAAC	GTGGAAATCCA	
5	64681	TCGAGGACGA	ACTCGCTGAG	GTACTGAAGG	GAGTTGCAGG	GAAGGCCGCG	TCGGTGGCGT
64741	GGTGGTCGAC	CGTGGACAGC	GCCTGGGTGA	CCGAGCCGGT	GGATGAGAGT	TACTGGTACC	
64801	GGAACCTGCG	TCGCCCCGTC	GCGCTGGACG	CGGCGGTGGC	GGAGCTGGAC	GGGTCCGTGT	
64861	TCGTGGAGTG	CAGCGCCCAT	CCGGTGCTGC	TGCCGGCGAT	GGAACAGGCC	CACACGGTGG	
64921	CGTCGTTGCG	CACCGGTGAC	GGCGGCTGGG	AGCGATGGCT	GACGGCGTTG	GCGCAGCGT	
10	64981	GGACCCCTGGG	CGCGGCAGTG	GACTGGGACA	CGGTGGTCGA	ACCGGTGCCA	GGCGGGCTGC
65041	TCGATCTGCC	CACCTACCGC	TTCGAGCGCC	GGCGCTACTG	GCTGGAAGCG	GCCGGTGCCA	
65101	CCGACCTGTC	CGCGGCCGGG	CTGACAGGGG	CAGCACATCC	CATGCTGGCC	GCCATCACGG	
65161	CACTACCCGC	CGACGACGGT	GGTGTGTTTC	TCACCGCCG	GATCTCGTTG	CGCACGCATC	
65221	CCTGGCTGGC	TGATCACCGC	GTGCGGGGCA	CGGTCTGCT	GCCCCGCACG	GCCTTTGTGG	
15	65281	AGCTGGTCAT	CCGGGCCGGT	GACGAGACCG	GTTGCGGAT	AGTGGATGAA	CTGGTCATCG
65341	AATCCCCCCT	CGTGGTGCCT	GCGACCGCAG	CCGTGGATCT	GTCGGTGACC	GTGGAAGGAG	
65401	CTGACGGAGGC	CGGACGGCGG	CGAGTGACCG	TCCACGCCG	CACCGAAGGC	ACCGGCAGCT	
65461	GGACCCGGCA	CGCCAGCGC	ACCCGTACCC	CCGACACCCC	CGACACCCCC	AACGCTTCCG	
20	65521	GTGTTGTCGG	TGCGGAGCCG	TTCTCGCAGT	GGCCACCTGC	CACTGCCGCG	GCGTCGACA
65581	CCTCGGAGTT	CTACTTGCGC	CTGGACGCGC	TGGGCTACCG	GTTCGGACCC	ATGTTCCCGC	
65641	GAATGCCGGC	TGCCTGGCGT	GATGGTACA	CCGTGTACGC	CGAGGTGCGC	CTCCCCGAGG	
65701	ACCGTGCCGC	CGACGCGGAC	GGTTTCGGCA	TGCACCCGGC	GCTGCTCGAC	GCGGCCTTGC	
65761	AGAGCGGCAG	CCTGCTCATG	CTGGAATCGG	ACGGCGAGCA	GAGCGTGCAA	CTGCCGTCT	
25	65821	CCTGGCACGG	CGTCCGGTTC	CACCGGACGG	GCGCGACCAT	GTCGGGGGTG	GCGGTGCGTAC
65881	CGGGCCTCGGA	CGGCCTCCGG	CTGCATGCCG	CGGACAGCGG	GAACCGTCCC	GTCGCGACGA	
65941	TCGACCGCGT	CGTGACCCGG	TCCCCCGGAAG	CGGACCTCGC	GCCCCCGCAT	CCGATGCTGC	
66001	GGGTCCGGTG	GGCCCCGGTG	CCCGTACCTG	CCGGGGCCGG	TCCGTCGAC	GCGGACCGTGC	
66061	TGACGCTGCG	CGCGGACGAC	GCCGACCCGC	TCGGGGAGAC	CCGGGACCTG	ACCACCCGTG	
66121	TTCTCGACGC	GCTGCTCCGG	GCCGACCGGC	CGGTGATCTT	CCAGGTGACC	GGTGGCCTCG	
30	66181	CCGCCAAGGC	GGCCGCGAGC	CTGGTCCGCA	CCGCTCAGAA	CGAGCAGCCC	GGCCGCTCT
66241	TCCCTCGTCGA	AACGGACCCG	GGAGAGGTCC	TGGACGGCGC	GAAGGCGGAC	GCGATCCGG	
66301	CACTCGGCGA	GCCCCATGTG	CGGCTCGCGC	ACGGCCCTTT	CGAGGCAGCC	CGGCTGATGC	
66361	GGGCCACGCC	GTCCCTGACG	CTCCCCGACA	CCGGGTCGTG	GCAGCTGC	CCGTCGGCCA	
66421	CCGGTCTCCCT	CGACGACCTT	GCCGTCGTCC	CCACCGACGC	CCCCGACCGG	CCGCTCGCG	
35	66481	CCGGCGAGGT	CGGGATCGCG	GTACCGCGG	CGGGCTGAA	CTTCCGGAT	GTCACGGTCG
66541	CGCTCGGTGT	GGTCGCCGAT	GCGCGTCCGC	TCGGCAGCGA	GGCCGCGGGT	GTCGTCTGG	
66601	AGACCGGCC	CGGTGTGCAC	GACCTGGCGC	CCGGCGACCG	GGTCCTGGGG	ATGCTCGCG	
66661	GCGCCTTCGG	ACCGGTCGCG	ATCACCGACC	GGCGGCTGCT	CGGCCGGATG	CCGGACGGCT	
66721	GGACGTTCCC	GCAGGCGGGG	TCCGTGATGA	CCGCGTTCGC	GACCGCGTGG	TACGGCCTGG	
40	66781	TCGACCTGGC	CGGGCTGCCG	CCCGCGGAGA	AGGTCTGAT	CCACCGGGCG	GCGACGGGTG
66841	TCGGCGCGGC	GGCCGTCCAG	ATCGCGCGC	ATCTGGCGC	GGAGGTGTAC	GCGACCA	
66901	GCGCCCGGAA	GCGCCATCTG	GTGGACCTGG	ACGGAGCGCA	TCTGGCCGAT	TCCCGCAGCA	
66961	CCGCGTTCGC	CGACGCGTTC	CCGCCGGTCG	ATGTCGTGCT	CAACTCGCTC	ACCGGTGAAT	
45	67021	TCCTCGACGC	GTCCGTGCGC	CTGCTCGCGG	CGGGTGGCG	GTTCATCGAG	ATGGGAAAGA
67081	CGGACATCCG	GCACGCCGTC	CAGCAGCCGT	TCGACCTGAT	GGACGCCGGC	CCCGACCGGA	
67141	TGCACTGGAT	CATCGTCGAG	CTGCTCGGCC	TGTCGCGCG	CGACGTGCTG	CACCGCTGC	
67201	CGGTCCACGC	CTGGGACGTC	CGGCAGGCCG	GGGAGGCCGT	CGGCTGGATG	AGCAGCGGGC	
67261	GTCACACCGG	CAAGCTGGTG	CTGACGGTCC	CGCGGCCGCT	GGATCCCGAG	GGGGCCGTG	
50	67321	TCATCACCGG	CGGCTCCGGC	ACCCCTGCCG	GCATCCTCGC	CCGCCACCTG	GGCCACCCCC
67381	ACACCTACCT	GCTCTCCCGC	ACCCCACCCC	CCGACACAC	CCCCGGCACC	CACCTCCCC	
67441	GCGACGTCGG	CGACCCCCAC	CAACTCGCCA	CCACCTCGC	CCGCATCCCC	CAACCCCTCA	
67501	CCGCCGTCTT	CCACACCGCC	GGAACCCCTCG	ACGACGCCCT	GTCGACAAC	CTCACCCCC	
67561	ACCGCGTCGA	CACCGTCCTC	AAACCCAAGG	CCGACGCCGC	CTGGCACCTG	CACCGGTCA	
67621	CCCGCGACAC	CGACCTCGCC	GGGTTGCTCG	TCTACTCCGC	GGTCGCCGGC	CTCATGGCA	
55	67681	GCCCCGGGCA	GGGCAACTAC	GTCGCGCGA	ACCGTCTCT	CGACCGCCTC	GCGAACACC
67741	GCCGTGCGCA	AGGGCTGCC	GCGCAGTCCC	TCGCATGGGG	CATGTGGCG	GACGTCA	
67801	CGCTCACCGC	GAAACTCAC	GACGCGGAC	GCCAGCGCAT	CCGGCGCAGC	GGATTCCC	
67861	CGTTGAGCGC	CGCGGACGCC	ATGCGGCTGT	TCGACGCCGC	GACCGTACCC	CCGGAACCGG	
67921	TCGTCGTCG	GACGACCGTC	GACCTCACCC	AGCTCGACGG	CGCCGTCGCG	CCGTTGCTCC	
60	67981	GCGGTCTGGC	CGCGCACCGG	GGCGGCCGG	CGCGCACGGT	CGCCGCCAAC	GCGGGCGAAG
68041	AGCCCCCTGGC	CGTGCCTCTT	GCCGGCGTGA	CCGCGGCCGA	GCAGCGGCC	ATCATGAGG	
68101	AGGTCTGCT	CCGCCACGCC	GCCGCGGTCC	TCGCGTACGG	GCTGGGCGAC	CGCGTGGCG	
68161	CGGACCGTCC	GTTCCGCGAG	CTCGGTTTCG	ATTGCTGAC	CGCGGTCGAC	CTGCGCAATC	
68221	GGCTCGCGGC	CGAGACGGGG	CTGCGGCTGC	CGACGACGCT	GGTGTTCAGC	CACCCGACGG	

	68281	CGGAGGGCGCT	CACCGCCCCAC	CTGCTCGAAC	TGATCGACGC	TCCCACCGCC	CGGATCGCCG
	68341	GGGAGTCCCT	GCCCCGGGTG	ACGGCCGCTC	CCGTGGCGGC	CGCGCGGGAC	CAGGACGAGC
5	68401	CGATCGCCAT	CGTGGCGATG	CGTGCGCCGC	TGCCC GGTTG	TGTGACGTCG	CCCGAGGACC
	68461	TGTGGGGCT	CGTCGAGTCC	GGCACCGACG	CGATCACCAAC	GCCTCCTGAC	GACCGCGGCT
	68521	GGGACGTCGA	CGCGCTGTAC	GACCGGGACC	CGGACGCCGC	CGGCAAGGCG	TACAACCTGC
	68581	GGGGCGGTAA	CCTGGCGGGG	CGGGCGGAGT	TCGACGCCGC	TTCTTCGAC	ATCAGTCCGC
	68641	GCGAAGCGCT	CGGCATGGAC	CCGCAGCAAC	GCCTGCTGCT	CGAAACGGCG	TGGGAGGCAG
10	68701	TCGAGCGCGG	CCGGATCAGT	CCGGCGTGC	TCCGCGGGCG	GGAGGTGCGC	GTCTATGTCG
	68761	GTGCGCCGC	GCAGGGCTAC	GGGCTGGCG	CGAGGGACAC	CGAGGGCCAC	GCGATCACCG
	68821	GTGGTCCAC	GAGCCTGCT	TCCGGACGGC	TGGCGTACGT	GCTGGGCTG	GAGGGCCCGG
	68881	CGGTCAACCGT	GGACACGGCG	TGCTCGTCG	CTCTGGTCG	GCTGATCTG	GCGTGC CAGG
15	68941	GGCTGCGCCT	GGGCGAGTGC	GAACTCGCTC	TGGCGGGAGG	GGTCCTCGTA	CTGAGTCGCG
	69001	CGGCCCGT	CGTGGAGTTC	TCCCGCCAGC	GGGGGCTCG	GGCCGACGGG	CGCTGCAAGT
	69061	CGTTCGGCGC	GGGCGCGGAC	GGCACGACGT	GGTCCGAGGG	CCTGGGCTG	CTCGTACTGG
	69121	AACGGCTCTC	CGACGCCGAG	CGGCTCGGGC	ACACCGTGC	CGCCGTCGTC	C CGGGCAGCG
	69181	CCGTCA CGTC	CGACGGCGCC	TCCAACGGCC	TCACCGCGCC	GAACGGGCTC	TCGCAGCAGC
	69241	GGGTCA TCCG	GAAGGCGCTC	GCCGCGGGCG	GGCTGACCGG	CGCCGACGTG	GACGTGTCG
	69301	AGGGGACCGG	CACCGGCACC	CGGCTCGGC	ACCCGGTCGA	GGCGGACCGG	CTGCTCGCGA
20	69361	CGTACGGGCA	GGACCGTCCG	GCACCGGTCT	GGCTGGGCTC	GCTGAAGTCG	AAACATCGGAC
	69421	ATGCCACGGC	CGCGGCCGGT	GTCGCGGGCG	TCATCAAGAT	GGTGCAGGCG	ATCGGCGCGG
	69481	GCACGATGCC	GGGGACGCTG	CATGTGGAGG	AGCCCTCGCC	CGCCGTCGAC	TGGAGCACCG
	69541	GACAGGTGTC	CCTGCTCGGC	TCCAACCGGC	CCTGGCCCGA	CGACGAGCGT	CCCGGCCGGG
	69601	CGGCCGTCTC	CGCGTT CGGG	CTCAGCGGG	CGAACCGCA	CGTCATCCTG	GAACAGCACC
25	69661	GTCCGGCGCC	CGTGGCGTCC	CAGCCGCCCC	GGCCGCCCCG	TGAGGAGTCC	CAGCCGCTGC
	69721	CGTGGGTGCT	CTCCCGCGG	ACTCCGGCCG	CGCTGCGGC	CCAGGCGGCC	CGGCTGCGCG
	69781	ACCACCTCGC	GGCGGCACCG	GACCGGGATC	CGTGGACAT	CGGGTACCGC	CTGGCCACCA
	69841	GCCGCGCCCA	GTTCGCCCCAC	CGTGCCGCGG	TCGTCGCCAC	CACCCCGGAC	GGATTCCGTG
	69901	CCGCGCTCGA	CGGCCCTCGCG	GACGGCGCGG	AGGCGCCCCG	AGTCGTCACC	GGGACCGCTC
	69961	AGGAGCGGCG	CGTCGCCCTC	CTCTTCGACG	GCCAGGGCGC	CCAGCGCGCC	GAATGGGGC
30	70021	GCGAGCTCCA	CCGCCGGTTC	CCCGCTTCG	CCGCCCGCTG	GGACGAGGTC	TCCGACCGT
	70081	TCGGCAAGCA	CCTCAAGCAC	TCCCCCACGG	ACGTCTACCA	CGCGAACAC	GGCGCTCTCG
	70141	CCCATGACAC	CCTGTACGCC	CAGGCCGGCC	TGTTCACGCT	CGAAGTGGCG	CTGCTCGGC
	70201	TGCTGGAGCA	CTGGGGGGTG	CGGCCGGACG	TGCTCGTCG	GCACTCCGTC	GGCGAGGTGA
	70261	CCGCGCGTA	CGCGCGGGGG	GTGCTCACCC	TGGCGGACGC	GACGGAGTTG	ATCGTGGCCC
35	70321	GGGGCGGGG	GCTCGGGCG	CTGCCGCCG	GGGCGATGCT	CGCCGTCGAC	GGAAAGCCGG
	70381	CGGAGGTGCG	CGCCCGCACG	GATCTGGACA	TCGCCGCCG	CAACGGCCCG	TCCGCCGTGG
	70441	TGCTGCCGT	CTCGCGGAC	GATGTGGCG	CGTTCGAACG	GGAGTGGTCG	GCGGCCGGGG
	70501	GGCGCACGAA	ACGGCTCGAC	GTCGGGCACG	CGTCCCACTC	CGGGCACGTC	GACGGTGC
	70561	TCGACGGCTT	CCGTACGGT	CTGGAGTCG	TCCGCTTCG	CGCCGCCGCG	CTGCCGGTGG
40	70621	TGTCACGAC	GACGGGCCGG	GACGCCGCGG	ACGACCTCAT	AACCCCCGCG	CACTGGCTGC
	70681	GCCATGCGCG	TCGGCCGGT	CTGTTCTCGG	ATGCCGTCG	GGAGCTGGCC	GACCGCGGCC
	70741	TCACCA CGTT	CGTGGCGCTC	GGCCCCCTCCG	GCTCCCTGGC	GTCGGCCGCG	GCGGAGAGCG
	70801	CGGGGGAGGA	CGCCGGGAC	TACACCGCGG	TGCTGCCGCG	CGGGACCGGT	GAGGAGACCG
	70861	CGGCGT GAC	CGCCCTCGCC	GAGCTGCACG	CCCACGGCGT	CCGGGTCGAC	CTGGCCCGGG
45	70921	TACTGGCGG	TGGCCGGCCA	GTGGACCTTC	CCGTGTACGC	GTTCCAGCAC	CGTTCC TACT
	70981	GGCTGCC	GGCCGTGGCG	GGGGCGCCGG	CCACCGTGGC	GGACACCGGG	GGTCCGGCGG
	71041	AGTCCGAGCC	GGAGGACCTC	ACCGTCGCCG	AGATCGTCCG	TCGGCGCACC	GCGGCCTGTC
	71101	TCGGCGTCAC	GGACCCCCGCC	GACGTGATG	CGGAAGCGAC	GTTCTCGCG	CTCGGTTTCG
	71161	ACTCACTGGC	GGTGCAGCGG	CTGCGCAACC	AGCTCGCC	GGCAACCGGG	CTGGACCTGC
50	71221	CGGCGCCGT	CCTGTTGAC	CACGACACCC	CGGCCGCCGT	CACCGCGTTC	CTCCAGGACC
	71281	GGATCGAGGC	CGGCCAGGAC	CGGATCGAGG	CGGGCGAGGA	CGACGACGCG	CCCACCGTGC
	71341	TCTCGCTCCT	GGAGGAGATG	GAGTCGCTCG	ACGCCGGGA	CATCGGGCG	ACGCCGGCCC
	71401	CGGAGCGTGC	GGCCATCGCC	GATCTGCTCG	ACAAGCTCGC	CCATACCTGG	AAGGACTACC
	71461	GATGAGCACC	GATA CGACG	AGGGAACGCC	GCCC GCCGGC	CGCTGCCCAT	TCGCGATCCA
55	71521	GGACGGTCAC	CGGCCATCC	TGGAGAGCGG	CACGGTGGGT	TCGTTCGACC	TGTTCGCGT
	71581	CAAGCACTGG	CTGGTCGCCG	CCGCCGAGGA	CGTCAAGCTG	GTCACCAACG	ATCCGCGGTT
	71641	CAGCTCGGCC	GGCCCGTCCG	AGATGCTG	CGACCGGGCG	CCGGGCTGGT	TCTCCGGGAT
	71701	GGACTCACCG	GAGCACAAAC	GCTACCGGC	GAAGATCGCG	GGGGACTTCA	CACTGCGCGC
	71761	GGCGCGCAAG	CGGGGAGGACT	TCGTCGCCG	GGCCGCCGAC	GCCTGCCCTGG	ACGACATCGA
60	71821	GGCCGCGGGA	CCC GGACCCG	ACCTCATCCC	CGGGTACGCC	AAGCGGCTGC	CCTCCCTCGT
	71881	CATCAACGCG	CTGTACGGGC	TCACCCCTGA	GGAGGGGGCC	GTGCTGGAGG	CACGGATGCG
	71941	CGACATCACC	GGCTCGGCCG	ATCTGGACAG	CGTCAAGACG	CTGACCGACG	ACTTCTTCGG
	72001	GCACGCGCTG	CGGCTGGTCC	GCGCGAAGCG	TGACGAGCGG	GGCGAGGACC	TGCTGCACCG
	72061	GCTGGCCTCG	GCCGACGACG	GCGAGATCTC	GCTCAGCGAC	GACGAGGCGA	CGGGCGTGT

72121 CGCGACGCTG CTGTCGCCG GCCACGACTC GGTGCAGCAG ATGGTCGGCT ACTGCCTCTA
 72181 CGCACTGCTC AGCCACCCCCG AGCAGCAGGC GGCGCTGCCG GCAGCGCCCG AGCTGGTCGA
 72241 CAACCGGGTC GAGGAGATGC TCCGTTCCCT GCCCCTAAC CAGATGGCG TACCGCGCGT
 72301 CTGTGTCGAG GACGTCGATG TGCAGGGCGT GCGCATCCGT GCAGGGCGACA ACGTGATCCC
 5 72361 GCTCTACTCG ACGGCCAACC GCGACCCCGA GGTGTTCCCG CAGCCCGACA CCTTCGATGT
 72421 GACGCCCGC CTGGAGGGCA ACTTCGCGTT CGGCCACGGC ATTACAAGT GTCCCAGCCA
 72481 GCACATCGCC CGGGTGTCA TCAAGGTGCG CTGCGTCCGG TTGTTGAGC GTTCCCCGA
 72541 CGTCCGGCTG GCCGGCGACG TGCGATGAA CGAGGGGCTC GGGCTGTTCA GCCCCGGCGA
 72601 GCTGCGGGTC ACCTGGGGGG CGGCATGAGT CACCCGGTGG AGACGTTGCG GTTGGCGAAC
 10 72661 GGGACGACGG TCGCGCACAT CAACCGGGC GAGGCGCAGT TCCTCTACCG GGAGATCTC
 72721 ACCCAGCGCT GCTACCTGCG CCACGGTGT GACCTGCGCC CGGGGGACGT GGTGTTGAC
 72781 GTCGGCGCGA ACATCGGCAT GTTCACGCTT TTCGCGCATC TGGAGTGTCC TGGTGTGACC
 72841 GTGCACGCT TCGAGCCCGC GCCCGTGCCG TTCGCGGCCG TGCGGGCGAA CGTGACCGGG
 72901 CACGGCATCC CGGGCCAGGC GGACCAAGTGC GCGGTCTCCG ACAGCTCCGG CACCCGGAAG
 15 72961 ATGACCTTCT ATCCCGACGC CACGCTGATG TCCGGTTTCC ACGCGGATGC CGCGGCCCGG
 73021 ACGGAGCTGT TGCGCACGCT CGGCCCTAAC GGCGGCTACA CGCCCGAGGA CGTCGACACC
 73081 ATGCTCGCGC AACTGCCCGA CGTCAGCGAG GAGATCGAAA CCCCTGTGGT CGGGCTCTCC
 73141 GACGTCATCG CGGAGCGCCG TATCGAGGCC ATCGGCCCTGC TGAAGGTCGA CGTGGAGAAG
 20 73201 AGCGAACGGC AGGTCTTCGC CGGCCTCGAG GACACCGACT GGCCCCGTAT CGGCCAGGTC
 73261 GTCGCGGAGG TCCACGACAT CGACGGCGCG CTCGAGGAGG TCGTCACGCT GCTCCGCGC
 73321 CATGGCTTCA CGTGGTTCGC CGAGCAGGAA CCGCTGTTCG CGGCCACGGG CATCCACCAAG
 73381 GTCGCCGCGC GGCGGGTGGC CGGCTGAGCG CGCTCGGGC CGCCGGCGTC CGCACCGGGCG
 73441 CGCCGGTGC CGACGGCGGC TCAGCCGGCG TCGGACAGTT CCTTGGGAGG TTGCTGACGG
 73501 CCCTTACCC CGAGCTTGC GAAACAGTTG GTGAGGTGCT GTTCCACCGT GCTGGAGGTG
 25 73561 ACGAACAGCT GGCTGGCGAT CTCCCTGTTG GTGCGCCCGA CGCGGCCGTG CGACGCCACC
 73621 CGCGCTCCG CCTCGGTCA CGATGTCGAT CGCTCGGCCG CGTCACGTC CTGGGTGCCG
 73681 TCCGCGTCCG AGGACTCCCC ACCGAGCCCG CGGAGGAGCG GCACGGCTCC GCACTGGGTC
 73741 GCGAGGTGCG GTGCGCGGC GAAACAGTCCC CGCGCACCGC TGTGCCCGCG GAGCATGCCG
 73801 CACGCTTCGC CCATGTCGGC GAGGACGCGG GCCAGCTCGT ACTGGTCGCG GCACATGATG
 30 73861 AGCAGATCGG CGGCCTCGTC GAGCAGTTG ATCCGTTGG CGGCCGGACT GTAGGCCGCC
 73921 TGCACCCGCA CGCTCATCAC CGCGCCCGG GACCCCATCG CGCCGGACAG CTGCTCGGAG
 73981 ATGAGCCTCA GCCCCCTCGC ACGGCCCGGG CGAGCAGCA GAAGCGCTTC GGCAGGCGTC
 74041 ACCCGCCACA GGGCCAGGGC CGGCACGTC ACGGACCAGC GTCCGATCCG CTCCCCCGAG
 74101 TCCCAGAACG CGTTGTACCG CGCCCGGTAC CGCCCGGGCG CGAGATGGTG TTGCCCCACGG
 35 74161 GCCCAGACCA TGTGAGTCC GAAAGAGGCTG TCGGAGGTCT CCTCCGGCAA CGGCTCGCG
 74221 AGCCACCGCT CGGCCCGGT CAGGTCGCCC AGTCGGATCG CGCGGCCAC GGTGCTGCTC
 74281 AGCGGCAATG CGGCGGCCAT CCCCCCAGGAG GGCACGACCC GGGGGGGCGAG CGCGGCCCTCG
 74341 CCGCATTGCA CGGCGGCCGGT CAGGTCGCCG CGGCGCAGCG CGGCCCTCGGC GCGGAACCCC
 74401 GCGTGGACCG CCTCGTCGGC CGGGGTCCGC ATGTTGTCGT CACCGGCCAG CTTGTCGACC
 40 74461 CAGGACTGGA CGGCATCGGT GTCCCGCGC TAGAGCAGGG CCAGCAACGC CATCATGGTC
 74521 GTGGTCCGGT CGTGTGAC CGGGAGTGC TGGAGCACGT ACTCGGCTTT GGCCTCGGCC
 74581 TGTTCGGACC AGCCGCGCAG CGCCTGCTC AGGGCCTTGT CGGGGACGGC GCGGTGCCGG
 74641 ACGGCTCCGG AAAACGAGGC GACCTCGTCC TCGGCCGGCG GATCGGCCGG ACGGGGGGGA
 74701 TCGGCCGCGC CGGGATAGAT CAGCGCGAGG GACAGGTCCG CGACGCGCAG GTGCGCCCGG
 45 74761 CCCTGTCGCG TCGGGCGGGC GGAGCGCTGG CGCCGCCAGGA CCTCGGCCGGC CTGCCCCGGC
 74821 CGCCCGTCCA TCGCCAGCCA GCAGGCGAGC GACACGGCGT GTCGCTGGA GAGGAGCCGT
 74881 TCCCAGCGACG CGGTGAGCAG CTCGGGCACA TGCCGGCCGG ATCTGGCGGG ATCGCAGAGC
 74941 CGCTCGATGG CGCGGGTGT GACGCGCAGT CGCGCGTGGA CGCGGGGGTC GTCGGAGGCC
 75001 CGGTAGGCGA ACTCCAGGT GGTGACGGCC TCGTCAGCT CGCCGCGCAG GTGGTGTCTG
 50 75061 CGCGCCCGT CGGTGAACAG CCCGGCGACC TCGCGCCCGT GCACCCGGCC GGTACCCATC
 75121 TGGTGGCGGG CGAGCACCT GCTGCCACG CGCGGGTCCC GCAGCAGTTC CAGGCCAGC
 75181 TCGTGCAGGC CACGCCGCTC GGCGCGGAG AGGTCGTCGA GTACGACGGA GCAGGCGCG
 75241 GGGTGGCGGG ACCGCCCTTC CGCGCAGCAG CGCCCGCTGA CCAGCTGTT GTGGGCCCTGC
 75301 TCGACCCGCT CGGTGTCGAG CGCGGTCA CGCTGGACGA GGGTGAGTTC GACACTCTCG
 55 75361 CCGAGCACGG CGGAAGCTCG CGCGACGCTC AGCGCGGGCG GGCCGCAACG ATAGAGCGAC
 75421 CCGAGGTAGG CGAGCCGGTA CGCCCGCCCC GCGACCACTT CCAGGCACCC TGAGGTCCTG
 75481 GTCCGTGCTT CCCGGATGTC GTCGATCAGG CGTGGCCGA GGAGCAGGTT CGCGCCGGTC
 75541 GCCCAGAACG CCTGGGCCAC CACGTCGTCG TGCGCTCTCGT GGCCGAGGT CGGGCGCACG
 75601 AGTCGGTGG TCTCGCCCTC GGTGAGCGGG CGCAGCGCGA TCTCCTGGTA GTGGCGCAGA
 60 75661 CTCAGCAGTG CGGCCCGGAA TTGGGAGTGG CGCGGGCGTCG GCCGGAGCAG CTCGGTCAGC
 75721 ACGATGGCGA CACGGGCCCG GCTGATGCGG CGCGCGAGGT GGAGCAGGCA CGCGCAGCGAC
 75781 GGCGCGTCGG CGTGGTGCAC GTCGTCGATG CGGATCAGTA CGGGCCGCTC CGCGGGCGAGC
 75841 GTCAGCACCG TGCGGGTGGAG TTCGGTCCCC AGGCGGTTGT CGACGTCGGC CGGCAGGTTT
 75901 TCGCACGATG CGTCAGCGC GACCAGCTCC GGTGTCCGGG CGGCCAGCTC GGGCTGGTCG

75961	AGGAGCTGGC	CGAGCATGCC	GTACGGCAGG	GCCCCGCTCCT	CCATGGAGCA	CACCGCGCGA	
76021	AGGGTGACGA	AGCCGGCCTT	GGCCCGCGCG	GCGTCGAGGA	GTTCGGTCTT	GCCGCAGGCG	
76081	ATCGGGCCGG	TGACGGCGC	GACGACGCC	CGCCCGCCCC	CCGCTCGGGT	GAGCGCCCGG	
76141	TGGAGGGAAC	CGAACTCGTC	ATCGCGGGCG	ATCAGGTCTG	GGGGAGATAA	GCGCGCTATC	
5	76201	ACGAATGGAA	CTACCTCGCG	ACCGTCGTGG	AAACCCATAG	GCATCACATG	GCTTGTGAT
	76261	CTGTACGGCT	GTGATTCAAGC	CTGGCGGGAT	GCTGTGCTAC	AGATGGGAAG	ATGTGATCTA
	76321	GGGCCGTGCC	GTTCCTCTAG	GAGCCGACCG	CCCCCGGCC	CACCCGCCGT	ACCCCCCTGGG
	76381	CCACCAGCTC	GGCGACCCCG	TCCTGGTGGT	CGACGAGGTA	GAAGTCCCCG	CCGGGAAAGA
	76441	CCTCCACCGT	GGTCGGCGCG	GTCGTGTGCC	CGGCCCAGGC	GTGGGCTGTC	TCCACCGTCG
10	76501	TCTTCGGATC	GTCGTCACCG	ATGCACACCG	TGATCGGGT	CTCCAGCGGC	GGCGCGGGCT
	76561	CCCACCGGTA	CGTCTCCGCC	GCGTAGTAGT	CCGCCCCGAA	CGGGGCCAGG	ATCAGCGCGC
	76621	GCATTTCGTC	GTCCGCCATC	ACATCGGC	TCGTCCCGCC	GAGGCCGATG	ACCGCCGCCA
	76681	GCAGCTCGTC	GTCGGACCGG	AGGTGGTCCCT	GGTCGGCGCG	CGGCTGCGAC	GGCGCCCGCC
15	76741	GGCCCAGAGAC	GATCAGGTGC	GCCACCGGGA	GCCGCTGGGC	CAGCTGAAC	GCGAGTGTG
	76801	CGCCCAGTGT	GTGGCCGAAC	AGCACCAGCG	GACGGTCCAG	CCCCGGCTTC	AACGCCCTCGG
	76861	CCACGAGGCC	GGCGAGAAACA	CGCAGGTCGC	GCACCGCTC	CTCGTCGCGG	CGGTCCCTGGC
	76921	GGCCGGGGTA	CTGCACGGCG	TACACGTCCG	CCACCGGGGC	GAGCGCACCG	GCCAGCGGAA
	76981	GGTAGAACGT	CGCCGATCCG	CCGGCGTGGG	GCAGCAGCAC	CACCCGTACC	GGGGCCTCGG
20	77041	GGGTGGGAA	GAAC TGCCGC	AGCCAGAGTT	CCGAGCTCAC	CGCACCCCCCT	CGGCCGCGAC
	77101	CTGGGGAGCC	CGGAACCGGG	TGATCTCGGC	CAAGTGCTTC	TCCCGCATCT	CGGGTCCG
	77161	CACGCCCAT	CCCTCCTCCG	GC GCCAGACA	GAGGACGCCG	ACTTTGCCGT	TGTGCACATT
	77221	GCGATGCACA	TCGCGCACCG	CCGACCCGAC	GTCGTCGAGC	GGTAGGTCA	CCGACAGCGT
	77281	CGGGTGCACC	ATCCCCCTTG	AGATCAGGCG	GTTGCCCTCC	CACGCCCTCAC	GATAGTTCGC
25	77341	GAAGTGGGTA	CCGATGATCC	GCTTCACCGGA	CATCCACAGG	TACCGATTGT	CAAAGGCGTG
	77401	CTCGTATCCC	GAGGTTGACG	CCGAGGTGAC	GATCGTGCCA	CCCCGACGTG	TCACGTAGAC
	77461	ACTCGCGCCG	AACGTCGCGC	GCCCCGGGTG	CTCGAACACAG	ATGTCGGGAT	CGTCACCGCC
	77521	GGTCAGCTCC	CGGATC				

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the FK-520 PKS of *Streptomyces hygroscopicus* is shown herein merely to illustrate a preferred embodiment of the invention, and the present invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences shown merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following general description of the FK-520 PKS genes and modules of the PKS proteins encoded thereby is provided. This general description is followed by a more detailed description of the various domains and modules of the FK-520 PKS contained in and encoded by the compounds of the invention. In this description, reference to a heterologous PKS refers to any PKS other than the FK-520 PKS. Unless otherwise indicated, reference to a PKS includes reference

to a portion of a PKS. Moreover, reference to a domain, module, or PKS includes reference to the nucleic acids encoding the same and vice-versa, because the methods and reagents of the invention provide or enable one to prepare proteins and the nucleic acids that encode them.

5 The FK-520 PKS is composed of three proteins encoded by three genes designated *fkbA*, *fkbB*, and *fkbC*. The *fkbA* ORF encodes extender modules 7 - 10 of the PKS. The *fkbB* ORF encodes the loading module (the CoA ligase) and extender modules 1 - 4 of the PKS. The *fkbC* ORF encodes extender modules 5 - 6 of the PKS. The *fkbP* ORF encodes the NRPS that attaches the pipecolic acid and cyclizes the FK-520
10 polyketide.

The loading module of the FK-520 PKS includes a CoA ligase, an ER domain, and an ACP domain. The starter building block or unit for FK-520 is believed to be a dihydroxycyclohexene carboxylic acid, which is derived from shikimate. The recombinant DNA compounds of the invention that encode the loading module of the
15 FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of methods and in a variety of compounds. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for the loading module of the
20 heterologous PKS is replaced by the coding sequence for the FK-520 loading module, provides a novel PKS coding sequence. Examples of heterologous PKS coding sequences include the rapamycin, FK-506, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the
25 coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, either replacing the CoA ligase with a different CoA
30 ligase, deleting the ER, or replacing the ER with a different ER. In addition, or alternatively, the ACP can be replaced by another ACP. In similar fashion, the corresponding domains in another loading or extender module can be replaced by one or more domains of the FK-520 PKS. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that
35 synthesizes FK-520, an FK-520 derivative, or another polyketide.

The first extender module of the FK-520 PKS includes a KS domain, an AT domain specific for methylmalonyl CoA, a DH domain, a KR domain, and an ACP domain. The recombinant DNA compounds of the invention that encode the first extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the FK-520 PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or only a portion of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting either the DH or KR or both; replacing the DH or KR or both with another DH or KR; and/or inserting an ER. In replacing or inserting KR, DH, and ER domains, it is often beneficial to replace the existing KR, DH, and ER domains with the complete set of domains desired from another module. Thus, if one desires to insert an ER domain, one may simply replace the existing KR and DH domains with a KR, DH, and ER set of domains from a module containing such domains. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a gene for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the first extender module of the FK-520 PKS.

In an illustrative embodiment of this aspect of the invention, the invention provides recombinant PKSs and recombinant DNA compounds and vectors that encode

such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and domains of an FK-520 or FK-520 derivative PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS
5 encoded thereby can be fed or supplied with N-acylcysteamine thioesters of novel precursor molecules to prepare FK-520 derivatives. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication Nos. US97/02358 and US99/03986, each of which is incorporated herein by reference.

The second extender module of the FK-520 PKS includes a KS, an AT specific
10 for methylmalonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the second extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 second extender module is inserted into a DNA compound that comprises the
15 coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the second
20 extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-
30 520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding
35

domains in a module of a heterologous PKS can be replaced by one or more domains of the second extender module of the FK-520 PKS.

The third extender module of the FK-520 PKS includes a KS, an AT specific for malonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of 5 the invention that encode the third extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence 10 for a module of the heterologous PKS is either replaced by that for the third extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding 15 sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 20 malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of 25 these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK- 30 520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the third extender module of the FK-520 PKS.

The fourth extender module of the FK-520 PKS includes a KS, an AT that binds ethylmalonyl CoA, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the fourth extender module of the FK-520 PKS and the 35 corresponding polypeptides encoded thereby are useful for a variety of applications. In

one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender 5 module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS 10 that produces an FK-520 derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the ethylmalonyl CoA specific AT with a malonyl CoA, methylmalonyl CoA, or 2-hydroxymalonyl CoA 15 specific AT; and/or deleting the inactive DH, inserting a KR, a KR and an active DH, or a KR, an active DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, a PKS for a polyketide other than FK-520, or from 20 chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fourth extender module of the FK-520 PKS.

As illustrative examples, the present invention provides recombinant genes, 25 vectors, and host cells that result from the conversion of the FK-506 PKS to an FK-520 PKS and vice-versa. In one embodiment, the invention provides a recombinant set of FK-506 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by 30 those for the AT domain of the fourth extender module of the FK-520 PKS. This recombinant PKS can be used to produce FK-520 in recombinant host cells. In another embodiment, the invention provides a recombinant set of FK-520 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of

the fourth extender module of the FK-506 PKS. This recombinant PKS can be used to produce FK-506 in recombinant host cells.

Other examples of hybrid PKS enzymes of the invention include those in which the AT domain of module 4 has been replaced with a malonyl specific AT domain to provide a PKS that produces 21-desethyl-FK520 or with a methylmalonyl specific AT domain to provide a PKS that produces 21-desethyl-21-methyl-FK520. Another hybrid PKS of the invention is prepared by replacing the AT and inactive KR domain of FK-520 extender module 4 with a methylmalonyl specific AT and an active KR domain, such as, for example, from module 2 of the DEBS or oleandolide PKS enzymes, to produce 21-desethyl-21-methyl-22-desoxo-22-hydroxy-FK520. The compounds produced by these hybrid PKS enzymes are neurotrophins.

The fifth extender module of the FK-520 PKS includes a KS, an AT that binds methylmalonyl CoA, a DH, a KR, and an ACP. The recombinant DNA compounds of the invention that encode the fifth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one or both of the DH and KR; replacing any one or both of the DH and KR with either a KR and/or DH; and/or inserting an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous fifth extender module coding sequence

can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fifth extender module of the FK-520 PKS.

- 5 In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the DH domain of the fifth extender module have been deleted or mutated to render the DH non-functional. In one such mutated gene, the KR and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. The resulting PKS genes code for the
- 10 expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-19 to C-20 double bond of FK-520 and has a C-20 hydroxyl group. Such analogs are preferred neurotrophins, because they have little or no immunosuppressant activity. This recombinant fifth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment,
- 15 the present invention provides a recombinant FK-520 PKS that contains both this fifth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that
- 20 express this recombinant PKS and so synthesize the corresponding (lacking the C-19 to C-20 double bond of FK-506 and having a C-20 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the DH domain of module 5 has been deleted or otherwise rendered inactive and thus produces this novel polyketide.
- 25 The sixth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the sixth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes
- 30 the FK-520 sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In
- 35 another embodiment, a DNA compound comprising a sequence that encodes the sixth

extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the sixth extender module coding sequence 5 is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the KR, DH, and ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, 10 DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous sixth extender module coding 15 sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the sixth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant 20 FK-520 PKS genes in which the coding sequences for the DH and ER domains of the sixth extender module have been deleted or mutated to render them non-functional. In one such mutated gene, the KR, ER, and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. This can also be accomplished by simply replacing the coding sequences for extender module six with those for an 25 extender module having a methylmalonyl specific AT and only a KR domain from a heterologous PKS gene, such as, for example, the coding sequences for extender module two encoded by the *eryAI* gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that has a C-18 hydroxyl group. Such 30 analogs are preferred neurotrophins, because they have little or no immunosuppressant activity. This recombinant sixth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this sixth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of 35 the FK-506 PKS. The invention also provides recombinant host cells derived from FK-

506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (having a C-18 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the DH and ER domains of module 6 have
5 been deleted or otherwise rendered inactive and thus produces this novel polyketide.

The seventh extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the seventh extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of
10 applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 seventh extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the seventh extender module of the FK-520 PKS or the latter is merely added to coding
15 sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the seventh extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

20 In another embodiment, a portion or all of the seventh extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or malonyl CoA specific AT; deleting the KR, the DH, and/or the ER; and/or replacing the
25 KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous
30 seventh extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the seventh extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the AT domain of the seventh extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes 5 code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-15 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such analogs are preferred, because they are more slowly metabolized than FK-520. This recombinant seventh extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an 10 illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this seventh extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK- 15 506 but that express this recombinant PKS and so synthesize the corresponding (C-15-desmethoxy) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 7 has been replaced and thus produces this novel polyketide.

In another illustrative embodiment, the present invention provides a hybrid PKS 20 in which the AT and KR domains of module 7 of the FK-520 PKS are replaced by a methylmalonyl specific AT domain and an inactive KR domain, such as, for example, the AT and KR domains of extender module 6 of the rapamycin PKS. The resulting hybrid PKS produces 15-desmethoxy-15-methyl-16-oxo-FK-520, a neurotrophin compound.

25 The eighth extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, and an ACP. The recombinant DNA compounds of the invention that encode the eighth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 30 eighth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the eighth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the eighth extender 35

module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

- In another embodiment, a portion of the eighth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or malonyl CoA specific AT; deleting or replacing the KR; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP.
- 5 In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous eighth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520
- 10 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the eighth extender module of the FK-520 PKS.
- 15

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the AT domain of the eighth extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-13 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such analogs are preferred, because they are more slowly metabolized than FK-520. This recombinant eighth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this eighth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (C-13-desmethoxy) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 8 has been replaced and thus produces this novel polyketide.

The ninth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the ninth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 ninth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the ninth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the ninth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the ninth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the KR, DH, and ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous ninth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the ninth extender module of the FK-520 PKS.

The tenth extender module of the FK-520 PKS includes a KS, an AT specific for malonyl CoA, and an ACP. The recombinant DNA compounds of the invention that encode the tenth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 tenth extender module is inserted into a DNA compound that comprises the coding sequence

for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the tenth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a 5 DNA compound comprising a sequence that encodes the tenth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion or all of the tenth extender module coding 10 sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; and/or inserting a KR, a KR and DH, or a KR, DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. 15 In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous tenth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that 20 synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the tenth extender module of the FK-520 PKS.

The FK-520 polyketide precursor produced by the action of the tenth extender module of the PKS is then attached to pipecolic acid and cyclized to form FK-520. The 25 enzyme FkbP is the NRPS like enzyme that catalyzes these reactions. FkbP also includes a thioesterase activity that cleaves the nascent FK-520 polyketide from the NRPS. The present invention provides recombinant DNA compounds that encode the *fkbP* gene and so provides recombinant methods for expressing the *fkbP* gene product in recombinant host cells. The recombinant *fkbP* genes of the invention include those in which the 30 coding sequence for the adenylation domain has been mutated or replaced with coding sequences from other NRPS like enzymes so that the resulting recombinant FkbP incorporates a moiety other than pipecolic acid. For the construction of host cells that do not naturally produce pipecolic acid, the present invention provides recombinant DNA compounds that express the enzymes that catalyze at least some of the biosynthesis of 35 pipecolic acid (see Nielsen *et al.*, 1991, *Biochem.* 30: 5789-96). The *fkbL* gene encodes a

homolog of RapL, a lysine cyclodeaminase responsible in part for producing the pipecolate unit added to the end of the polyketide chain. The *fkbB* and *fkbL* recombinant genes of the invention can be used in heterologous hosts to produce compounds such as FK-520 or, in conjunction with other PKS or NRPS genes, to produce known or novel polyketides and non-ribosomal peptides.

The present invention also provides recombinant DNA compounds that encode the P450 oxidase and methyltransferase genes involved in the biosynthesis of FK-520. Figure 2 shows the various sites on the FK-520 polyketide core structure at which these enzymes act. By providing these genes in recombinant form, the present invention provides recombinant host cells that can produce FK-520. This is accomplished by introducing the recombinant PKS, P450 oxidase, and methyltransferase genes into a heterologous host cell. In a preferred embodiment, the heterologous host cell is *Streptomyces coelicolor* CH999 or *Streptomyces lividans* K4-114, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. In addition, by providing recombinant host cells that express only a subset of these genes, the present invention provides methods for making FK-520 precursor compounds not readily obtainable by other means.

In a related aspect, the present invention provides recombinant DNA compounds and vectors that are useful in generating, by homologous recombination, recombinant host cells that produce FK-520 precursor compounds. In this aspect of the invention, a native host cell that produces FK-520 is transformed with a vector (such as an SCP2* derived vector for *Streptomyces* host cells) that encodes one or more disrupted genes (i.e., a hydroxylase, a methyltransferase, or both) or merely flanking regions from those genes. When the vector integrates by homologous recombination, the native, functional gene is deleted or replaced by the non-functional recombinant gene, and the resulting host cell thus produces an FK-520 precursor. Such host cells can also be complemented by introduction of a modified form of the deleted or mutated non-functional gene to produce a novel compound.

In one important embodiment, the present invention provides a hybrid PKS and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the present invention a hybrid PKS is a recombinant PKS that comprises all or part of one or more modules and thioesterase/cyclase domain of a first PKS and all or part of one or more modules, loading module, and thioesterase/cyclase

domain of a second PKS. In one preferred embodiment, the first PKS is all or part of the FK-520 PKS, and the second PKS is only a portion or all of a non-FK-520 PKS.

One example of the preferred embodiment is an FK-520 PKS in which the AT domain of module 8, which specifies a hydroxymalonyl CoA and from which the C-13 methoxy group of FK-520 is derived, is replaced by an AT domain that specifies a malonyl, methylmalonyl, or ethylmalonyl CoA. Examples of such replacement AT domains include the AT domains from modules 3, 12, and 13 of the rapamycin PKS and from modules 1 and 2 of the erythromycin PKS. Such replacements, conducted at the level of the gene for the PKS, are illustrated in the examples below. Another 10 illustrative example of such a hybrid PKS includes an FK-520 PKS in which the natural loading module has been replaced with a loading module of another PKS. Another example of such a hybrid PKS is an FK-520 PKS in which the AT domain of module three is replaced with an AT domain that binds methylmalonyl CoA.

In another preferred embodiment, the first PKS is most but not all of a non-FK-15 520 PKS, and the second PKS is only a portion or all of the FK-520 PKS. An illustrative example of such a hybrid PKS includes an erythromycin PKS in which an AT specific for methylmalonyl CoA is replaced with an AT from the FK-520 PKS specific for malonyl CoA.

Those of skill in the art will recognize that all or part of either the first or second 20 PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See U.S. provisional patent application Serial No. 60/091,526, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or 25 domain. For purposes of the present invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:

(i) from fusions of heterologous domain (where heterologous means the domains 30 in that module are from at least two different naturally occurring modules) coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS,
but also:

(ii) from fusions of heterologous module (where heterologous module means two 35 modules are adjacent to one another that are not adjacent to one another in naturally

occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,

(iii) from expression of one or more FK-520 PKS genes with one or more non-FK-520 PKS genes, including both naturally occurring and recombinant non-FK-520

5 PKS genes, and

(iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described herein.

Examples of the production of a hybrid PKS by co-expression of PKS genes from 10 the FK-520 PKS and another non-FK-520 PKS include hybrid PKS enzymes produced by coexpression of FK-520 and rapamycin PKS genes. Preferably, such hybrid PKS enzymes are produced in recombinant *Streptomyces* host cells that produce FK-520 or FK-506 but have been mutated to inactivate the gene whose function is to be replaced by the rapamycin PKS gene introduced to produce the hybrid PKS. Particular examples 15 include (i) replacement of the *fkbC* gene with the *rapB* gene; and (ii) replacement of the *fkbA* gene with the *rapC* gene. The latter hybrid PKS produces 13,15-didesmethoxy-FK-520, if the host cell is an FK-520 producing host cell, and 13,15-didesmethoxy-FK-506, if the host cell is an FK-506 producing host cell. The compounds produced by these hybrid PKS enzymes are immunosuppressants and neurotrophins but can be readily 20 modified to act only as neurotrophins, as described in Example 6, below.

Other illustrative hybrid PKS enzymes of the invention are prepared by replacing the *fkbA* gene of an FK-520 or FK-506 producing host cell with a hybrid *fkbA* gene in which: (a) the extender module 8 through 10, inclusive, coding sequences have been replaced by the coding sequences for extender modules 12 to 14, inclusive, of the 25 rapamycin PKS; and (b) the module 8 coding sequences have been replaced by the module 8 coding sequence of the rifamycin PKS. When expressed with the other, naturally occurring FK-520 or FK-506 PKS genes and the genes of the modification enzymes, the resulting hybrid PKS enzymes produce, respectively, (a) 13-desmethoxy-FK-520 or 13-desmethoxy-FK-506; and (b) 13-desmethoxy-13-methyl-FK-520 or 13-30 desmethoxy-13-methyl-FK-506. In a preferred embodiment, these recombinant PKS genes of the invention are introduced into the producing host cell by a vector such as pHU204, which is a plasmid pRM5 derivative that has the well-characterized SCP2* replicon, the *coleI* replicon, the *tsr* and *bla* resistance genes, and a *cos* site. This vector can be used to introduce the recombinant *fkbA* replacement gene in an FK-520 or FK-35 506 producing host cell (or a host cell derived therefrom in which the endogenous *fkbA*

gene has either been rendered inactive by mutation, deletion or homologous recombination with the gene that replaces it) to produce the desired hybrid PKS.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units," *Biochemistry* 38(5):1643-1651, incorporated herein by reference. Stereochemistry can also be changed by changing the KR domain. Also, one can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases," *Science* 284: 482-485, incorporated herein by reference.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the present invention.

Avermectin

U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemalectin.

MacNeil *et al.*, 1992, *Gene* 115: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

Ikeda *et al.*, Aug. 1999, Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*, *Proc. Natl. Acad. Sci. USA* 96: 9509-9514.

Candidin (FR008)

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

Epothilone

U.S. Pat. App. Serial No. 60/130,560, filed 22 April 1999.

Erythromycin

5 PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large
multifunctional polypeptide in the erythromycin producing polyketide synthase of
10 *Saccharopolyspora erythraea*.

Glycosylation Enzymes

PCT Pat. App. Pub. No. 97/23630 to Abbott.

FK-506

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of
15 the immunosuppressant FK-506, *Eur. J. biochem.* 256: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide
synthase involved in the biosynthesis of the macrolide immunosuppressant FK-506, *Eur.*
J. Biochem. 244: 74-80.

Methyltransferase

20 US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from
Streptomyces MA6858. 31-O-desmethyl-FK-506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and
hydroxylase genes involved in the biosynthesis of the immunosuppressants FK-506 and
FK-520, *J. Bacteriol.* 178: 5243-5248.

25 ***Streptomyces hygroscopicus***

U.S. patent application Serial No. 09/154,083, filed 16 Sep. 1998.

Lovastatin

U.S. Pat. No. 5,744,350 to Merck.

Narbomycin

30 U.S. patent application Serial No. 60/107,093, filed 5 Nov. 1998, and Serial No.
60/120,254, filed 16 Feb. 1999.

Nemadectin

MacNeil *et al.*, 1993, *supra*.

Niddamycin

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelstis*, *J. Bacteriol.* 179: 7515-7522.

Oleandomycin

- Swan *et al.*, 1994, Characterisation of a *Streptomyces antibioticus* gene encoding 5 a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

U.S. patent application Serial No. 60/120,254, filed 16 Feb. 1999.

- Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases 10 responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

Picromycin

PCT patent application US99/15047, filed 2 Jul. 1999.

- Xue *et al.*, 1998, Hydroxylation of macrolactones YC-17 and narbomycin is 15 mediated by the *pikC*-encoded cytochrome P450 in *Streptomyces venezuelae*, *Chemistry & Biology* 5(11): 661-667.

Xue *et al.*, Oct. 1998, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity, *Proc. Natl. Acad. Sci. USA* 95: 12111 12116.

20 **Platenolide**

EP Pat. App. Pub. No. 791,656 to Lilly.

Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

- 25 Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

Rifamycin

- August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: 30 deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S669, *Chemistry & Biology*, 5(2): 69-79.

Sorangium PKS

U.S. patent application Serial No. 09/144,085, filed 31 Aug. 1998.

Soraphen

- 35 U.S. Pat. No. 5,716,849 to Novartis.

Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

5 **Spiramycin**

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

Tylosin

10 EP Pub. No. 791,655 to Lilly.

U.S. Pat. No. 5,876,991 to Lilly.

Kuhstoss *et al.*, 1996, *Gene* 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

Tailoring enzymes

15 Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of polyketide synthase genes that serve as readily available sources of DNA and sequence information for use in 20 constructing the hybrid PKS-encoding DNA compounds of the invention. Methods for constructing hybrid PKS-encoding DNA compounds are described without reference to the FK-520 PKS in PCT patent publication No. 98/51695; U.S. Patent Nos. 5,672,491 and 5,712,146 and U.S. patent application Serial Nos. 09/073,538, filed 6 May 1998, and 09/141,908, filed 28 Aug 1998, each of which is incorporated herein by reference.

25 The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Moreover, there are often two or more modules in the hybrid PKS in which all or part of the module is derived from a second (or third) PKS. Thus, as one illustrative example, the present invention provides a hybrid FK-520 PKS that contains the naturally occurring loading module and FkbP as well as modules 30 one, two, four, six, seven, and eight, nine, and ten of the FK-520 PKS and further contains hybrid or heterologous modules three and five. Hybrid or heterologous module three contains an AT domain that is specific of methylmalonyl CoA and can be derived for example, from the erythromycin or rapamycin PKS genes. Hybrid or heterologous module five contains an AT domain that is specific for malonyl CoA and can be derived 35 for example, from the picromycin or rapamycin PKS genes.

While an important embodiment of the present invention relates to hybrid PKS enzymes and corresponding genes, the present invention also provides recombinant FK-520 PKS genes in which there is no second PKS gene sequence present but which differ from the FK-520 PKS gene by one or more deletions. The deletions can encompass one or more modules and/or can be limited to a partial deletion within one or more modules. When a deletion encompasses an entire module, the resulting FK-520 derivative is at least two carbons shorter than the gene from which it was derived. When a deletion is within a module, the deletion typically encompasses a KR, DH, or ER domain, or both DH and ER domains, or both KR and DH domains, or all three KR, DH, and ER domains.

To construct a hybrid PKS or FK-520 derivative PKS gene of the invention, one can employ a technique, described in PCT Pub. No. 98/27203 and U.S. patent application Serial No. 08/989,332, filed 11 Dec. 1997, each of which is incorporated herein by reference, in which the large PKS gene is divided into two or more, typically three, segments, and each segment is placed on a separate expression vector. In this manner, each of the segments of the gene can be altered, and various altered segments can be combined in a single host cell to provide a recombinant PKS gene of the invention. This technique makes more efficient the construction of large libraries of recombinant PKS genes, vectors for expressing those genes, and host cells comprising those vectors.

Thus, in one important embodiment, the recombinant DNA compounds of the invention are expression vectors. As used herein, the term expression vector refers to any nucleic acid that can be introduced into a host cell or cell-free transcription and translation medium. An expression vector can be maintained stably or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a gene that serves to produce RNA that is translated into a polypeptide in the cell or cell extract. Furthermore, expression vectors typically contain additional functional elements, such as resistance-conferring genes to act as selectable markers.

The various components of an expression vector can vary widely, depending on the intended use of the vector. In particular, the components depend on the host cell(s) in which the vector will be used or is intended to function. Vector components for expression and maintenance of vectors in *E. coli* are widely known and commercially available, as are vector components for other commonly used organisms, such as yeast cells and *Streptomyces* cells.

In a preferred embodiment, the expression vectors of the invention are used to construct recombinant *Streptomyces* host cells that express a recombinant PKS of the invention. Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which do not produce

- 5 actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference.

The present invention provides a wide variety of expression vectors for use in 10 *Streptomyces*. For replicating vectors, the origin of replication can be, for example and without limitation, a low copy number vector, such as SCP2* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), 15 SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and SG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, 20 *Plasmid* 30: 131-140, each of which is incorporated herein by reference). Generally, however, high copy number vectors are not preferred for expression of genes contained 25 on large segments of DNA. For non-replicating and integrating vectors, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and pBR. For phage based vectors, the phages phiC31 and KC515 can be employed (see Hopwood *et al.*, *supra*).

Typically, the expression vector will comprise one or more marker genes by 30 which host cells containing the vector can be identified and/or selected. Useful antibiotic resistance conferring genes for use in *Streptomyces* host cells include the *ermE* (confers resistance to erythromycin and other macrolides and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4* (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes.

The recombinant PKS gene on the vector will be under the control of a promoter, 35 typically with an attendant ribosome binding site sequence. The present invention

provides the endogenous promoters of the FK-520 PKS and related biosynthetic genes in recombinant form, and these promoters are preferred for use in the native hosts and in heterologous hosts in which the promoters function. A preferred promoter of the invention is the *fkbO* gene promoter, comprised in a sequence of about 270 bp between 5 the start of the open reading frames of the *fkbO* and *fkbB* genes. The *fkbO* promoter is believed to be bi-directional in that it promotes transcription of the genes *fkbO*, *fkbP*, and *fkbA* in one direction and *fkbB*, *fkbC*, and *fkbL* in the other. Thus, in one aspect, the present invention provides a recombinant expression vector comprising the promoter of the *fkbO* gene of an FK-520 producing organism positioned to transcribe a gene other 10 than *fkbO*. In a preferred embodiment the transcribed gene is an FK-520 PKS gene. In another preferred embodiment, the transcribed gene is a gene that encodes a protein comprised in a hybrid PKS.

Heterologous promoters can also be employed and are preferred for use in host 15 cells in which the endogenous FK-520 PKS gene promoters do not function or function poorly. A preferred heterologous promoter is the *actI* promoter and its attendant activator gene *actII-ORF4*, which is provided in the pRM1 and pRM5 expression vectors, *supra*. This promoter is activated in the stationary phase of growth when secondary metabolites are normally synthesized. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melCI* gene, which act constitutively, and 20 the *tipA* gene and the *merA* gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is 25 placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actII-ORF4* gene discussed above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, *supra*) to activate promoters under their control.

30 In addition to providing recombinant DNA compounds that encode the FK-520 PKS, the present invention also provides DNA compounds that encode the ethylmalonyl CoA and 2-hydroxymalonyl CoA utilized in the synthesis of FK-520. Thus, the present invention also provides recombinant host cells that express the genes required for the biosynthesis of ethylmalonyl CoA and 2-hydroxymalonyl CoA. Figures 3 and 4 show the

location of these genes on the cosmids of the invention and the biosynthetic pathway that produces ethylmalonyl CoA.

For 2-hydroxymalonyl CoA biosynthesis, the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes are sufficient to confer this ability on *Streptomyces* host cells. For conversion of 2-hydroxymalonyl to 2-methoxymalonyl, the *fkbG* gene is also employed. While the complete coding sequence for *fkbH* is provided on the cosmids of the invention, the sequence for this gene provided herein may be missing a T residue, based on a comparison made with a similar gene cloned from the ansamitocin gene cluster by Dr. H. Floss. Where the sequence herein shows one T, there may be two, resulting in an extension of the *fkbH* reading frame to encode the amino acid sequence:

MTIVKCLVWLDLNTLWRGTVLEDDEVLTDEIREVITLDDRGILQAVASKNDH
DLAWERLERLGVAEYFVLARIGWGPKSQSVREIATELNFAPTTIAFIDDQPAERA
EVAFHLPEVRCPYPAEQAAATLLSLPEFSPPVSTVDSRRRLMYQAGFARDQAREA
YSGPDEDFLRSLDLSMTIAPAGEEEELSRVEELTLRTSQMNATGVHYSADLRAL
15 LTDPAHEVLVVTMGDRFGPHGAVGIILLEKKPSTWHLKLLATCRVVSFGAGAT
ILNWLTQGARAGAHLVADFRRTDRNRMMEIAYRFAGFADSDCPCVSEVAGAS
AAGVERLHLEPSARPAPTTLTAADIAPVTVSAAG.

For ethylmalonyl CoA biosynthesis, one requires only a crotonyl CoA reductase, which can be supplied by the host cell but can also be supplied by recombinant expression of the *fkbS* gene of the present invention. To increase yield of ethylmalonyl CoA, one can also express the *fkbE* and *fkbU* genes as well. While such production can be achieved using only the recombinant genes above, one can also achieve such production by placing into the recombinant host cell a large segment of the DNA provided by the cosmids of the invention. Thus, for 2-hydroxymalonyl and 2-methoxymalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the left side of the FK-520 PKS genes shown in Figure 1. For ethylmalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the right side of the FK-520 PKS genes shown in Figure 1 or, alternatively, both the right and left segments of DNA.

30 The recombinant DNA expression vectors that encode these genes can be used to construct recombinant host cells that can make these important polyketide building blocks from cells that otherwise are unable to produce them. For example, *Streptomyces coelicolor* and *Streptomyces lividans* do not synthesize ethylmalonyl CoA or 2-hydroxymalonyl CoA. The invention provides methods and vectors for constructing recombinant *Streptomyces coelicolor* and *Streptomyces lividans* that are able to

synthesize either or both ethylmalonyl CoA and 2-hydroxymalonyl CoA. These host cells are thus able to make polyketides, those requiring these substrates, that cannot otherwise be made in such cells.

- In a preferred embodiment, the present invention provides recombinant 5 *Streptomyces* host cells, such as *S. coelicolor* and *S. lividans*, that have been transformed with a recombinant vector of the invention that codes for the expression of the ethylmalonyl CoA biosynthetic genes. The resulting host cells produce ethylmalonyl CoA and so are preferred host cells for the production of polyketides produced by PKS enzymes that comprise one or more AT domains specific for ethylmalonyl CoA.
- 10 Illustrative PKS enzymes of this type include the FK-520 PKS and a recombinant PKS in which one or more AT domains is specific for ethylmalonyl CoA.

In a related embodiment, the present invention provides *Streptomyces* host cells in which one or more of the ethylmalonyl or 2-hydroxymalonyl biosynthetic genes have been deleted by homologous recombination or rendered inactive by mutation. For 15 example, deletion or inactivation of the *fkbG* gene can prevent formation of the methoxyl groups at C-13 and C-15 of FK-520 (or, in the corresponding FK-506 producing cell, FK-506), leading to the production of 13,15-didesmethoxy-13,15-dihydroxy-FK-520 (or, in the corresponding FK-506 producing cell, 13,15-didesmethoxy-13,15-dihydroxy-FK-506). If the *fkbG* gene product acts on 2-hydroxymalonyl and the resulting 2-methoxymaionyl substrate is required for incorporation by the PKS, the AT domains of 20 modules 7 and 8 may bind malonyl CoA and methylmalonyl CoA. Such incorporation results in the production of a mixture of polyketides in which the methoxy groups at C-13 and C-15 of FK-520 (or FK-506) are replaced by either hydrogen or methyl.

This possibility of non-specific binding results from the construction of a hybrid 25 PKS of the invention in which the AT domain of module 8 of the FK-520 PKS replaced the AT domain of module 6 of DEBS. The resulting PKS produced, in *Streptomyces lividans*, 6-dEB and 2-desmethyl-6-dEB, indicating that the AT domain of module 8 of the FK-520 PKS could bind malonyl CoA and methylmalonyl CoA substrates. Thus, one could possibly also prepare the 13,15-didesmethoxy-FK-520 and corresponding FK-506 30 compounds of the invention by deleting or otherwise inactivating one or more or all of the genes required for 2-hydroxymalonyl CoA biosynthesis, i.e., the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes. In any event, the deletion or inactivation of one or more biosynthetic genes required for ethylmalonyl and/or 2-hydroxymalonyl production prevents the formation of polyketides requiring ethylmalonyl and/or 2-hydroxymalonyl for biosynthesis, and the

resulting host cells are thus preferred for production of polyketides that do not require the same.

The host cells of the invention can be grown and fermented under conditions known in the art for other purposes to produce the compounds of the invention. See, e.g., 5 U.S. Patent Nos. 5,194,378; 5,116,756; and 5,494,820, incorporated herein by reference, for suitable fermentation processes. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. Preferred compounds of the invention include the following compounds: 13-desmethoxy-FK-506; 13-desmethoxy-FK-520; 13,15-didesmethoxy-FK-506; 13,15-didesmethoxy-FK-520; 13-desmethoxy-18-hydroxy-FK-506; 13-desmethoxy-18-hydroxy-FK-520; 13,15-didesmethoxy-18-hydroxy-FK-506; and 13,15-didesmethoxy-18-hydroxy-FK-520. These compounds can be further modified as described for tacrolimus and FK-520 in U.S. Patent Nos. 5,225,403; 5,189,042; 5,164,495; 5,068,323; 10 4,980,466; and 4,920,218, incorporated herein by reference.

15 Other compounds of the invention are shown in Figure 8, Parts A and B. In Figure 8, Part A, illustrative C-32-substituted compounds of the invention are shown in two columns under the heading R. The substituted compounds are preferred for topical administration and are applied to the dermis for treatment of conditions such as psoriasis. In Figure 8, Part B, illustrative reaction schemes for making the compounds shown in 20 Figure 8, Part A, are provided. In the upper scheme in Figure 8, Part B, the C-32 substitution is a tetrazole moiety, illustrative of the groups shown in the left column under R in Figure 8, Part A. In the lower scheme in Figure 8, Part B, the C-32 substitution is a disubstituted amino group, where R₃ and R₄ can be any group similar to the illustrative groups shown attached to the amine in the right column under R in Figure 25 8, Part A. While Figure 8 shows the C-32-substituted compounds in which the C-15-methoxy is present, the invention includes these C-32-substituted compounds in which C-15 is ethyl, methyl, or hydrogen. Also, while C-21 is shown as substituted with ethyl or allyl, the compounds of the invention includes the C-32-substituted compounds in which C-21 is substituted with hydrogen or methyl.

30 To make these C-32-substituted compounds, Figure 8, Part B, provides illustrative reaction schemes. Thus, a selective reaction of the starting compound (see Figure 8, Part B, for an illustrative starting compound) with trifluoromethanesulfonic anhydride in the presence of a base yields the C-32 O-triflate derivative, as shown in the upper scheme of Figure 8, Part B. Displacement of the triflate with 1H-tetrazole or 35 triazole derivatives provides the C-32 tetrazole or triazole derivative. As shown in the

lower scheme of Figure 8, Part B, reacting the starting compound with p-nitrophenylchloroformate yields the corresponding carbonate, which, upon displacement with an amino compound, provides the corresponding carbamate derivative.

The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation contains one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use. Suitable formulation processes and compositions for the compounds of the present invention are described with respect to tacrolimus in U.S. Patent Nos. 5,939,427; 5,922,729; 5,385,907; 5,338,684; and 5,260,301, incorporated herein by reference. Many of the compounds of the invention contain one or more chiral centers, and all of the stereoisomers are included within the scope of the invention, as pure compounds as well as mixtures of stereoisomers. Thus the compounds of the invention may be supplied as a mixture of stereoisomers in any proportion.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases relating to immunosuppression or neuronal damage, a compound of the invention may be administered orally, topically,

parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

5 Dosage levels of the compounds of the present invention are of the order from about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the
10 present invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral
15 administration to humans may contain from 0.5 mg to 5 g of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention can be formulated within the range of,
20 for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight. The compounds and compositions of the invention are useful in treating disease conditions using doses and administration schedules as described for tacrolimus in U.S. Patent Nos. 5,542,436; 5,365,948; 5,348,966; and 5,196,437, incorporated herein by reference. The compounds
25 of the invention can be used as single therapeutic agents or in combination with other therapeutic agents. Drugs that can be usefully combined with compounds of the invention include one or more immunosuppressant agents such as rapamycin, cyclosporin A, FK-506, or one or more neurotrophic agents.

It will be understood, however, that the specific dosage level for any particular
30 patient will depend on a variety of factors. These factors include the activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the present invention and shall not be construed as being a limitation on the scope of the invention or claims.

5

Example 1

Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-520

The C-13 methoxyl group is introduced into FK-520 via an AT domain in extender module 8 of the PKS that is specific for hydroxymalonyl and by methylation of the hydroxyl group by an S-adenosyl methionine (SAM) dependent methyltransferase.

10 Metabolism of FK-506 and FK-520 primarily involves oxidation at the C-13 position into an inactive derivative that is further degraded by host P450 and other enzymes. The present invention provides compounds related in structure to FK-506 and FK-520 that do not contain the C-13 methoxy group and exhibit greater stability and a longer half-life *in vivo*. These compounds are useful medicaments due to their immunosuppressive and 15 neurotrophic activities, and the invention provides the compounds in purified form and as pharmaceutical compositions.

The present invention also provides the novel PKS enzymes that produce these novel compounds as well as the expression vectors and host cells that produce the novel PKS enzymes. The novel PKS enzymes include, among others, those that contain an AT 20 domain specific for either malonyl CoA or methylmalonyl CoA in module 8 of the FK-506 and FK-520 PKS. This example describes the construction of recombinant DNA compounds that encode the novel FK-520 PKS enzymes and the transformation of host cells with those recombinant DNA compounds to produce the novel PKS enzymes and the polyketides produced thereby.

25 To construct an expression cassette for performing module 8 AT domain replacements in the FK-520 PKS, a 4.6 kb *Sph*I fragment from the FK-520 gene cluster was cloned into plasmid pLitmus 38 (a cloning vector available from New England Biolabs). The 4.6 kb *Sph*I fragment, which encodes the ACP domain of module 7 followed by module 8 through the KR domain, was isolated from an agarose gel after 30 digesting the cosmid pKOS65-C31 with *Sph* I. The clone having the insert oriented so the single *Sac*I site was nearest to the *Spe*I end of the polylinker was identified and designated as plasmid pKOS60-21-67. To generate appropriate cloning sites, two linkers were ligated sequentially as follows. First, a linker was ligated between the *Spe*I and *Sac*I sites to introduce a *Bgl*II site at the 5' end of the cassette, to eliminate interfering 35 polylinker sites, and to reduce the total insert size to 4.5 kb (the limit of the phage

KC515). The ligation reactions contained 5 picomolar unphosphorylated linker DNA and 0.1 picomolar vector DNA, i.e., a 50-fold molar excess of linker to vector. The linker had the following sequence:

5 5'-CTAGTGGGCAGATCTGGCAGCT-3'
3'-ACCCGTCTAGACCG-5'

The resulting plasmid was designated pKOS60-27-1.

Next, a linker of the following sequence was ligated between the unique *Sph*I and *Af*III sites of plasmid pKOS60-27-1 to introduce an *Nsi*I site at the 3' end of the module 8 cassette. The linker employed was:

10 5'-GGGATGCATGGC-3'
3'-GTACCCCTACGTACCGAATT-5'

The resulting plasmid was designated pKOS60-29-55.

To allow in-frame insertions of alternative AT domains, sites were engineered at the 5' end (*Avr* II or *Nhe* I) and 3' end (*Xho* I) of the AT domain using the polymerase chain reaction (PCR) as follows. Plasmid pKOS60-29-55 was used as a template for the PCR and sequence 5' to the AT domain was amplified with the primers SpeBgl-fwd and either *Avr*-rev or *Nhe*-rev:

15 SpeBgl-fwd 5'-CGACTCACTAGTGGGCAGATCTGG-3'
Avr-rev 5'-CACGCCTAGGCCGGTCGGTCTCGGGCCAC-3'
20 *Nhe*-rev 5'-GCGGCTAGCTGCTCGCCCATCGCGGGATGC-3'

The PCR included, in a 50 µl reaction, 5 µl of 10x *Pfu* polymerase buffer (Stratagene), 5 µl 10x z-dNTP mixture (2 mM dATP, 2 mM dCTP, 2 mM dTTP, 1 mM dGTP, 1 mM 7-deaza-GTP), 5 µl DMSO, 2 µl of each primer (10 µM), 1 µl of template DNA (0.1 µg/µl), and 1 µl of cloned *Pfu* polymerase (Stratagene). The PCR conditions were 95°C for 2 min., 25 cycles at 95°C for 30 sec., 60°C for 30 sec., and 72°C for 4 min., followed by 4 min. at 72°C and a hold at 0°C. The amplified DNA products and the Litmus vectors were cut with the appropriate restriction enzymes (*Bgl*II and *Avr*II or *Spe*I and *Nhe*I), and cloned into either pLitmus 28 or pLitmus38 (New England BioLabs), respectively, to generate the constructs designated pKOS60-37-4 and pKOS60-37-2, respectively.

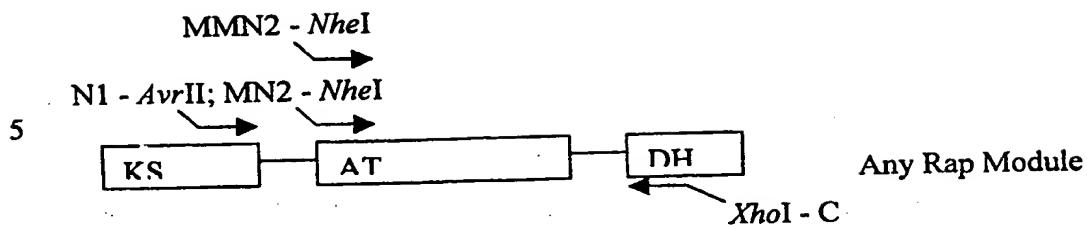
30 Plasmid pKOS60-29-55 was again used as a template for PCR to amplify sequence 3' to the AT domain using the primers BsrXho-fwd and *Nsi*Afl-rev:

35 BsrXho-fwd 5'-GATGTACAGCTCGAGTCGGCACGCCGCCGCATC-3'
*Nsi*Afl-rev 5'-CGACTCACTTAAGCCATGCATCC-3'
PCR conditions were as described above. The PCR fragment was cut with *Bsr*GI and *Af*III, gel isolated, and ligated into pKOS60-37-4 cut with *Asp*718 and *Af*III and

inserted into pKOS60-37-2 cut with *Bsr*GI and *Afl*II, to give the plasmids pKOS60-39-1 and pKOS60-39-13, respectively. These two plasmids can be digested with *Avr*II and *Xba*I or *Nhe*I and *Xba*I, respectively, to insert heterologous AT domains specific for malonyl, methylmalonyl, ethylmalonyl, or other extender units.

5 Malonyl and methylmalonyl-specific AT domains were cloned from the rapamycin cluster using PCR amplification with a pair of primers that introduce an *Avr*II or *Nhe*I site at the 5' end and an *Xba*I site at the 3' end. The PCR conditions were as given above and the primer sequences were as follows:

- 10 RATN1 5'-ATCCTAGGCGGGCRGGYGTGTCGTCCCTCGG-3'
(3' end of Rap KS sequence and universal for malonyl and methylmalonyl CoA),
RATMN2 5'-ATGCTAGCCGCCGCGTCCCCGTCTCGCGCG-3'
(Rap AT shorter version 5'- sequence and specific for malonyl CoA),
RATMMN2 5'-ATGCTAGCGGATTCGTCGGTGGTGTTCGCCGA-3'
15 (Rap AT shorter version 5'- sequence and specific for methylmalonyl CoA), and
RATC 5'-ATCTCGAGCCAGTASCGCTGGTGYTGGAAAGG-3'
(Rap DH 5'- sequence and universal for malonyl and methylmalonyl CoA).



Because of the high sequence similarity in each module of the rapamycin cluster, each primer was expected to prime any of the AT domains. PCR products representing ATs specific for malonyl or methylmalonyl extenders were identified by sequencing individual cloned PCR products. Sequencing also confirmed that the chosen clones contained no cloning artifacts. Examples of hybrid modules with the rapamycin AT12 and AT13 domains are shown in a separate figure.

The *AvrII-Xhol* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below. The AT of rap module 12 is specific for incorporation of malonyl units.

20	AGATCTGGCAGCTCGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50 I W Q L A E A L L T L V R E S T
	GCCGCCGTGCTCGGCCACGTGGTGGCGAGGACATCCCCGCACGGCGGC 100 A A V L G H V G G E D I P A T A A
	GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCCAGCTCGCAACG 150 F K D L G I D S L T A V Q L R N
25	CCCTCACCGAGGCAGCCGGTGTGCGGCTAACGCCACGGCGGTCTTCGAC 200 A L T E A T G V R L N A T A V F D
	TTCCCGACCCCGCACGTGCTCGCCGGAAAGCTCGGCACGAACTGACCGG 250 F P T P H V L A G K L G D E L T G
30	CACCCCGCGCGCCCGTCGTGCCCGGACCGCGGCCACGGCGGTGCGCACG 300 T R A P V V P R T A A T A G A H
	ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGGGTC 350 D E P L A I V G M A C R L P G G V
	GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400 A S P E E L W H L V A S G T D A I
35	CACGGAGTCCCGACGGACCGCGGCTGGGACGTCGACGCCATACGACC 450 T E F P T D R G W D V D A I Y D
	CGGACCCCGACGCGATCGGCAAGACCTTCGTCGGCACGGTGGCTTCCTC 500 P D P D A I G K T F V R H G G F L
40	ACCGGCGCGACAGGCTTCGACGCCGGCTTCGGCATCAGCCCGCGCA 550 T G A T G F D A A F F G I S P R E
	GGCCCTCGCGATGGACCCCGCAGCAGCGGGTGCTCCTGGAGACGTCGTGGG 600 A L A M D P Q Q R V L L E T S W
45	AGGCCTTCGAAAGCGCCGGCATCACCCCGGACTCGACCCGCCAGCGAC 650 T G V F V G A F S Y G Y G T G A D
	ACCCGGCGTGTTCGTCGGCGCCTCTCCTACGGTTACGGCACCGGTGCGGA 700 CACCGACGGCTTCGCGCGACCCGGCTCGCAGACCAGTGTGCTCTCCGGCC 750
50	T D G F G A T G S Q T S V L S G GGCTGTCTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800 R L S Y F Y G L E G P A V T V D T GCGTGTTCGTCGTGGTGGCGCTGCACCAGGCCGGCAGTCGCTGCG 850 A C S S S L V A L H Q A G Q S L R

CTCCGGCGAATGCTCGCTGCCCTGGTCGGCGCGTCACGGTGATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCGGCGGCTTCGTGGAGTTCTCCGGCAGCGCGGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 5 GGCGGGCGAAGGCCTTCGGCGGGTGCAGGACGGCACGAGCTTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 GGGTGCCTGGTGTGCTGAGAGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCACACCCTGGCGGTCTGGTCCGTCGGTCAACCCAGGATGGT 1100
 10 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGCTGTCGGCGCCAACGGGCCCTCGCAGGAGC 1150
 A S N G L S A P N G P S Q E R V I
 CGGGCAGGCCCTGGCAAACGCCGGCTCACCCGGCGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 15 TCGAGGCCACGGCACCGGACCGAGCTGGCGACCCCACATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCCACCTACGGACAGGAGCGCGCACCCCCCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGCTCCGGCGTC 1350
 20 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGGAGCTGCCGCCACG 1400
 G I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGAGCCGTCGCCGACGTCGACTGGACGCCGGCGCGT 1450
 L H A D E P S P H V D W T A G A V
 25 CGAACTGCTGACGTGGCCCGGCCGTGGCCCGAGACCGACCGCCCTAGGC 1500
 E L L T S A R P W P E T D R P R
 GGGCAGGCGTGTGTCGTCCTCGGGATCAGTGGCACCAACGCCACGTCA 1550
 R A G V S S F G I S G T N A H V I
 CTGGAAAGCGCACCCCCCACTCAGCCTGCCGACGGTGTAGCGAGCG 1600
 30 L E S A P P T Q P A D N A V I E R
 GGCACCGGAGTGGGTGCCGTTGGTATTCGGCCAGGACCCAGTCGGCTT 1650
 A P E W V P L V I S A R T Q S A
 TGACTGAGCACGAGGGCCGGTGTGGTGTGCTATCTGGCGCGTCGCCGG 1700
 L T E H E G R L R A Y L A A S P G
 35 GTGGATATGCCGGCTGTGGCATCGACGCTGGCGATGACACGGTCGGT 1750
 V D M R A V A S T L A M T R S V F
 CGAGCACCGTGCCGTGCTGGAGATGACACCGTCACCGGACCGCTG 1800
 E H R A V L L G D D T V T G T A
 TGTCTGACCTCGGGCGGTGTCGTCCTCCCCGGACAGGGGTGCGACCGT 1850
 40 V S D P R A V F V F P G Q G S Q R
 GCTGGCATGGGTGAGGAACTGGCCGCCGCGTCCCCGTCTCGCGCGGAT 1900
 A G M G E E L A A A F P V F A R I
 CCATCAGCAGGTGTGGGACCTGCTCGATGTGCCGATCTGGAGGTGAACG 1950
 H Q Q V W D L L D V P D L E V N
 45 AGACCGGTTACGCCACGGCCCTGTCGCAATGCAGGTGGCTCTGTT 2000
 E T G Y A Q P A L F A M Q V A L F
 GGGCTGCTGGAATCGTGGGTGTACGACCGGACGCCGGTGTGATCGGCATT 2050
 G L L E S W G V R P D A V I G H S
 GGTGGGTGAGCTGCGGCTCGTGTGTCGGGTGTGGTGTGGAGG 2100
 50 V G E L A A A Y V S G V W S L E
 ATGCCCTGCACCTTGGTGTGCGCGGGCTCGTCTGATGCAGGCTCTGCC 2150
 D A C T L V S A R A R L M Q A L P
 GCGGGTGGGGTGTGGTGTGCTGCCGGTCTCGGAGGGATGAGGCCGGC 2200
 A G G V M V A V P V S E D E A R A
 55 CGTGTGGGTGAGGGTGTGGAGATGCCCGCGTCAACGCCCGTCTGG 2250
 V L G E G V E I A A V N G P S S
 TGGTTCTCTCCGGTGATGAGGCCGGTGTGCTGCAGGCCGGAGGGCTG 2300
 V V L S G D E A A V L Q A A E G L
 GGGAAAGTGGACGCCGGTGGCGACCCAGCCACGCCGGTCCATTCCGCC 2350
 60 G K W T R L A T S H A F H S A R M
 GGAACCCATGCTGGAGGGAGTTCCGGCGGTGCGCAAGGCCCTGACCTACC 2400
 E P M L E E F R A V A E G L T Y
 GGACGCCGCAGGTCTCCATGGCCGGTGTGGTGTGACGGTGACCAACCGCTGAG 2450
 R T P Q V S M A V G D Q V T T A E

TACTGGGTGCCGCAGGTCCGGGACACGGTCCGGTCGGCGAGCAGGTGGC 2500
 Y W V R Q V R D T V R F G E Q V A
 CTCGTACGAGGACGCCGTGTCGAGCTGGGTGCCGACCGGTCACTGG 2550
 S Y E D A V F V E L G A D R S L
 5 CCCGCCTGGTCGACGGTGTGCGATGCTGCACGGCGACCGAAATCCAG 2600
 A R L V D G V A M L H G D H E I Q
 GCCGCATGGCGCCCTGGCCCACCTGTATGTCACGGCGTCACGGTCGA 2650
 A A I G A L A H L Y V N G V T V D
 CTGGCCCGCCTGGCGATGCTCCGGCACACGGGTGCTGGACCTTC 2700
 10 W P A L L G D A P A T R V L D L
 CGACATACGCCCTCCAGCACCGCCTACTGGCTCGAGTCGGCACGCCCG 2750
 P T Y A F Q H Q R Y W L E S A R P
 GCCGCATCCGACGGGCCACCCCGTGCCTGGCTCCGGTATGCCCTCGC 2800
 A A S D A G H P V L G S G I A L A
 15 CGGGTCGCCGGGCCGGTGTTCACGGGTTCCGTGCCGACCGGTGCGGACC 2850
 G S P G R V F T G S V P T G A D
 GCGCGGTGTCGCGAGCTGGCCTGGCGCCGGACGGTCGAC 2900
 R A V F V A E L A L A A A D A V D
 TGCGCCACGGTCGAGCGGCTCGACATGCCCTCCGTGCCCGGCCGGGG 2950
 20 C A T V E R L D I A S V P G R P G
 CCATGGCCGACGACCGTACAGACCTGGGTGACGAGCCGGCGACGACG 3000
 H G R T T V Q T W V D E P A D D
 GCCGGCGCCGGTTCACCGTGACACCCGACCGGGCGACGCCCGTGGACG 3050
 25 G R R R F T V H T R T G D A P W T
 CTGCACGCCGAGGGGGTGTGCGCCCCCATGGCACGCCCTGCCGATGC 3100
 L H A E G V L R P H G T A L P D A
 GGCGACGCCGAGTGGCCCCCACCGGGCGCGGTGCCCGGGACGGCTGC 3150
 A D A E W P P P G A V P A D G L
 CGGGTGTGGCGCCGGGGGACCAAGGTCTCGCCGAGGCCGAGGTGGAC 3200
 30 P G V W R R G D Q V F A E A E V D
 GGACCGGACGGTTTCGTGGTGACCCCGACCTGCTCGACGCCGGTCTTC 3250
 G P D G F V V H P D L L D A V F S
 CGCGGTGGCGACGGAAGCCGCCAGCCGGATGGCGCGACCTGACGG 3300
 A V G D G S R Q P A G W R D L T
 35 TGCACCGCTGGACGCCACCGTACTGCCCTGCCCTACCCGGCGACC 3350
 V H A S D A T V L R A C L T R R T
 GACGGAGCCATGGGATTGCCGCTTCGACGGCGCCGGCTGCCGGTACT 3400
 D G A M G F A A F D G A G L P V L
 CACCGCGGAGGGCGGTGACGCTGCCGGAGGTGGCGTACCGTCCGGCTCG 3450
 40 T A E A V T L R E V A S P S G S
 AGGAGTCGGACGGCTGACCGGTTGGAGTGGCTCGCGGTGCCGAGGCG 3500
 E E S D G L H R L E W L A V A E A
 GTCTACGACGGTGACCTGCCGAGGGACATGTCTGATACCGCCGCCA 3550
 V Y D G D L P E G H V L I T A A H
 45 CCCCGACGACCCCGAGGGACATACCCACCCGCCACACCCGCCACCC
 P D D P E D I P T R A H T R A T
 GCGCTCTGACCGCCCTGCAACACCCACCTCACCACCGACCCCTC 3650
 R V L T A L Q H H L T T T D H T L
 ATCGTCCACACCACCGACCCGCCGGCGCCACCGTACCGCCCTCAC 3700
 50 I V H T T T D P A G A T V T G L T
 CCGCACCGCCAGAACGAACACCCACCGCATCCGCTCATCGAAACCG 3750
 R T A Q N E H P H R I R L I E T
 ACCACCCCCACACCCCCCTCCCCCTGGCCAACTCGCCACCCCTCGACCAC 3800
 55 D H P H T P L P L A Q L A T L D H
 CCCCACCTCCGCCCTACCCACCCACACCCCTCCACCAACCCCCACCTCACCC 3850
 P H L R L T H H T L H H P H L T P
 CCTCCACACCACCCACCCACCCACCCACCAACCCCCCTCAACCCCGAACACG 3900
 L H T T T P P T T T P L N P E H
 CCATCATCATCACCGCGGCTCCGGCACCCCTCGCCGGCATCCTGCCCGC 3950
 60 A I I I T G G S G T L A G I L A R
 CACCTGAACCACCCCCACACCTACCTCCCTCCCGCACCCACCCCCCGA 4000
 H L N H P H T Y L L S R T P P P D
 CGCCACCCCCGGCACCCACCTCCCTCGCACGTCGGCGACCCCCACCAAC 4050
 A T P G T H L P C D V G D P H Q

TCGCCACCACCCCTACCCACATCCCCAACCCCTCACGCCATCTTCCAC 4100
 L A T T L T H I P Q P L T A I F H
 ACCGGCAGCCACCCCTCGACGGACGGCATCCTCCACGCCCTCACCCCGACCG 4150
 T A A T L D D G I L H A L T P D R
 5 CCTCACCACCGTCCCTCCACCCAAAGCCAACGCCGCTGGCACCTGCACC 4200
 L T T V L H P K A N A A A W H L H
 ACCTCACCCAAAACCAACCCCTCACCCACTTCGTCCCTACTCCAGCGCC 4250
 H L T Q N Q P L T H F V L Y S S A
 GCCGCCGTCCCTCGGCAGCCCCGACAAGGAAACTACGCCGCCAACGC 4300
 10 A A V L G S P G Q G N Y A A A A N A
 CTTCTCGACGCCCTCGCCACCCACCGCCACACCCCTCGGCCAACCGCCA 4350
 F L D A L A T H R H T L G Q P A
 CCTCCATCGCCATGGGCATGTGGCACACCACAGCACCCCTCACCGGACAA 4400
 15 T S I A W G M W H T T S T L T G Q
 CTCGACGACGCCGACGGGACCGCATCCGCCGGCGGTTCTCCCGAT 4450
 L D D A D R D R I R R G G F L P I
 CACGGACGACGAGGGCATGGGATGCAT
 T D D E G

20 The *AvrII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 13 (specific for methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

AGATCTGGCAGCTCGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 25 Q L A E A L L T L V R E S T
 GCCGGCGTGCTCGGCCACGTGGGTGGCGAGGACATCCCCCGGACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCCAGCTGCGCAACG 150
 F K D L G I D S L T A V Q L R N
 30 CCCTCACCGAGGCAGCGACGGGTGTGCGGCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGAAAGCTCGGCACGAACTGACCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCGCGCGCCCGTGTGCCCCGGACCGCGGCCACGGCGGTGCCACG 300
 35 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGACCGACGCCAT 400
 A S P E E L W H L V A S G T D A I
 40 CACGGAGTTCCCGACGGACCGCGGCTGGACGTCGACCGATCTACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACCGCGATCGGCAAGACCTTCGTCGGCAGGGTGGCTCCCTC 500
 P D P D A I G K T F V R H G G F L
 45 ACCGGCGCGACAGGGCTTCGACGCCGGTCTCGGCATCAGCCCGCGCA 550
 T G A T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCCGCAGCAGCGGGTGTCTGGAGACGTGCGTGGG 600
 A L A M D P Q Q R V L L E T S W
 AGGCCTTCGAAAGCGCCGGCATACCCCGGACTCGACCCCGGGCAGCGAC 650
 50 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCTTCTCCTACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGACCGGCTCGCAGACCGAGCTGTGCTCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 55 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTCGTCGGCTGGCGCTGCACCCAGGCGGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTGCCCTGGTCGGCGGTACGGTGATGGCGT 900
 S G E C S L A L V G G V T V M A
 60 CTCCGGCGGCTTCGTTCTCCGGCAGCGGGCTCGGCCGGAC 950

S P G G F V E F S R Q R G L A P D
 GGCGGGCGAAGGCCTTCGGCGCGGGTGCACGGCACGAGCTTCGGGA 1000
 G R A K A F G A G A D G T S F A E
 GGGTGCCTGGTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 5 G A G V L I V E R L S D A E R N
 GTCACACCCTGGCGGTGCTGGCGGTCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGCTGTGCGGCCAACGGCCGTCGAGGAGCAGGGTGAT 1150
 A S N G L S A P N G P S Q E R V I
 10 CGCGCAGGCCCTGGCCAACGCCGGCTACCCCGCGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCACCTACGGACAGGAGCGCCACCCCCCTGCTGCTGGG 1300
 15 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGTCGGCGTCGCCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGAGCTGCCGCCAG 1400
 G I I K M V Q A L R H G E L P P T
 20 CTGCACGCCGACGAGCCGTGCCGACGTCGACTGGACGCCGGCGCCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTGGCCCGGCCGTGGCCCGAGACCGACCGCCCTAGGC 1500
 E L L T S A R P W P E T D R P R
 GGGCGGGCGTGTGCTCCTCGGAGTCAGCGCACCAACGCCACGTAC 1550
 25 R A G V S S F G V S G T N A H V I
 CTGGAGAGCCACCCCCCGCTCAGCCCGGGAGGAGGCGCAGCCTGTTGA 1600
 L E S A P P A Q P A E E A Q P V E
 GACGCCGGTGGCTGGATGTGCTGCCCTGGTGTATATCGGCCAAGA 1650
 T P V V A S D V L P L V I S A K
 30 CCCAGCCCCTGACCGAACACGAAGACGGCTGCCGCCAACCTGGCG 1700
 T Q P A L T E H E D R L R A Y L A
 GCGTCGCCCGGGCGGATATACGGGCTGTGGCATCGACGCTGGCGGTGAC 1750
 A S P G A D I R A V A S T L A V T
 ACGGTGGTGTGAGCACCGCCGTAACCTGGAGATGACACCGTCA 1800
 35 R S V F E H R A V L L G D D T V
 CCGGCACCGCGGTGACCGACCCCAGGATCGTGTGTTGCTTCCGGCAG 1850
 T G T A V T D P R I V F V F P G Q
 GGGTGGCAGTGGCTGGGATGGCAGTGCAGTCGCGATTGTCGGTGG 1900
 G W Q W L G M G S A L R D S S V V
 40 GTTCGCCAGCGGATGGCGAGTGTGCGGCCGGTGCAGGAGTTCGTGG 1950
 F A E R M A E C A A A A L R E F V
 ACTGGGATCTGTCACGGTCTGGATGATCCGGCGGTGGGACCGGGTT 2000
 D W D L F T V L D D P A V V D R V
 GATGTGGTCCAGCCGCTCCCTGGCGATGATGGTTCCCTGGCGCGGT 2050
 45 D V V Q P A S W A M M V S L A A V
 GTGGCAGGCCGGTGTGCGGCCGATGCGGTGATGCCATTGCGAGG 2100
 W Q A A G V R P D A V I G H S Q
 GTGAGATCGCCGAGCTTGTGGCGGGTGCAGTCAGCGATGCC 2150
 G E I A A A C V A G A V S L R D A
 50 GCCCGGATCGTGACCTTGGCGAGCCAGGCGATGCCGGGCTGGCGG 2200
 A R I V T L R S Q A I A R G L A G
 CCGGGCGCGATGGCATCCGTCGCCCTGCCCGCGCAGGATGCGAGCTGG 2250
 R G A M A S V A L P A Q D V E L
 TCGACGGGGCTGGATGCCGCCAACACGGGGCGCTCCACCGTGCATC 2300
 55 V D G A W I A A H N G P A S T V I
 GCGGGCACCCCCGGAGCGGTGACCATGTCCTACCGCTCATGAGGCACA 2350
 A G T P E A V D H V L T A H E A Q
 AGGGGTGGGGTGCAGGCGATACCGTCGACTATGCCCTGCACACCCGC 2400
 G V R V R R I T V D Y A S H T P
 60 ACGTCGAGCTGATCCCGACGAACACTCGACATCACTAGCGACAGCAGC 2450
 H V E L I R D E L L D I T S D S S
 TCGCAGACCCCGCTGTGCCGTGGCTGTCGACCGTGGACGGCACCTGGGT 2500
 S Q T P L V P W L S T V D G T W V
 CGACAGCCCGTGGACGGGAGTACTGGTACCGAACCTGCGTAACCGG 2550

D S P L D G E Y W Y R N L R E P
 TCGGTTCCACCCCGCCGTCAAGCCAGTTGCAGGCCAGGGCAGACCCGTG 2600
 V G F H P A V S Q L Q A Q G D T V
 TTTCGAGGTCAAGGCCAGCCGGTGTGTCAGGCATGGACGACGA 2650
 5 F V E V S A S P V L L Q A M D D D
 TGTGTCACGGTGCCACCGCTGCGTGTGACGACGGCAGGCCACCCGGA 2700
 V V T V A T L R R D D G D A T R
 TGCTCACGCCCTGGCACAGGCCTATGTCCACGGCGTACCGTCGACTGG 2750
 M L T A L A Q A Y V H G V T V D W
 10 CCCGCCATCCTCGGACCAACCAACCCGGTACTGGACCTTCCGACCTA 2800
 P A I L G T T T R V L D L P T Y
 CGCCTCCAACACCAGCGTACTGGCTCGAGTCGGCACGCCGGCGCAT 2850
 A F Q H Q R Y W L E S A R P A A
 CCGACGCCGGCACCCGTGCTGGCTCCGGTATGCCCTCGCCGGT 2900
 15 S D A G H P V L G S G I A L A G S
 CCGGGCCGGGTGTTCACGGGTTCCGTGCGACCCGGTGCAGGCCGGT 2950
 P G R V F T G S V P T G A D R A V
 GTTCGTCGCCGAGCTGGCGCTGGCCGCCGGACGCGGTGACTGCCA 3000
 F V A E L A L A A A D A V D C A
 20 CGGTCGAGCGGCTCGACATGCCCTCCGTGCCCGGCCGGCATGGC 3050
 T V E R L D I A S V P G R P G H G
 CGGACGACCGTACAGACCTGGTCGACGAGCCGGGACGACGCCGGCG 3100
 R T T V Q T W V D E P A D D G R R
 CCGGTTCACCGTGCACACCCGCACCGCGACGCCCGTGGACGCTGACG 3150
 25 R F T V H T R T G D A P W T L H
 CCGAGGGGGTGCTGCGCCCCCATGGCACGGCCCTGCCGATGCCGCG 3200
 A E G V L R P H G T A L P D A A D
 GCCGAGTGGCCCCCACCGGGCGGGTGCCTGCCGCGAGGGCTGCCGGT 3250
 30 A E W P P P G A V P A D G L P G V
 GTGGCGCCGGGGGACCAAGGTCTCGCCGAGGCGAGGTGGACGGACCG 3300
 W R R G D Q V F A E A E V D G P
 ACGGTTTCGTGGTGCACCCGACCTGCTCGACGCCGGTCTCTCCGCGTC 3350
 D G F V V H P D L L D A V F S A V
 GGCGACGGAAGCCGCCAGCCGGGATGGCGCACCTGACGGTGCACGC 3400
 35 G D G S R Q P A G W R D L T V H A
 GTCGGACGCCACCGTACTGCGCCCTGCCACCCGGCGACCGACGGAG 3450
 S D A T V L R A C L T R R T D G
 CCATGGGATTGCGCCCTTCGACGGCCGGCTGCCGGTACTCACCGCG 3500
 40 A M G F A A F D G A G L P V L T A
 GAGGCCGGTACGCTGCCGGAGGTGGCGTCACCGTCCGGCTCCGAGGAGTC 3550
 E A V T L R E V A S P S G S E E S
 GGACGCCCTGCACCGGTTGGAGTGGCTCGCGTCGCCGAGGCCGGTACG 3600
 D G L H R L E W L A V A E A V Y
 ACGGTGACCTGCCGAGGGACATGTCTGATCACCGCCGCCACCCGAC 3650
 45 D G D L P E G H V L I T A A H P D
 GACCCCGAGGACATACCCACCCGCCACACCCGCCACCCCGTCCCT 3700
 D P E D I P T R A H T R A T R V L
 GACCGCCCTGCAACACCACTCACCACCCGACACCCCTCATCGTCC 3750
 T A L Q H H L T T D H T L I V
 50 ACACCAACCCGACCCGCCGGCGCCACCGTCACCGCCCTCACCGCACC 3800
 H T T T D P A G A T V T G L T R T
 GCCCAGAACGAAACACCCGCCACCGCATCCGCTCATCGAAACCGACCA 3850
 A Q N E H P H R I R L I E T D H P
 CCACACCCCCCTCCCCCTGGCCCAACTGCCACCCCGACCCCGACC 3900
 55 H T P L P L A Q L A T L D H P H
 TCCGCCTCACCCACCAACCCCTCCACCAACCCCACTCACCCCGTCCAC 3950
 L R L T H H T L H H P H L T P L H
 ACCACCAACCCACCCACCAACCCGCCACCCCGAACACGCCATCAT 4000
 60 T T T P P T T T P L N P E H A I I
 CATCACCGGGCGGCTCCGGCACCCCTGCCGGCATCTGCCGCCACCTGA 4050
 I T G G S G T L A G I L A R H L
 ACCACCCCCACACCTACCTCTCCCGCACCCACCCCGACGCCACC 4100
 N H P H T Y L L S R T P P P D A T
 CCCGGCACCCACCTCCCCCTGCCGACGTGGCGACCCCCACCAACTGCCAC 4150

P G T H L P C D V G D P H Q L A T
 CACCCCTCACCCACATCCCCAACCCCTCACGCCATCTTCCACACCGCCG 4200
 T L T H I P Q P L T A I F H T A
 CCACCCCTCGACGACGGCATCCTCCACGCCCTACCCCCGACCGCCCTCAC 4250
 5 A T L D D G I L H A L T P D R L T
 ACCGTCTCCACCCCACAGCAACGCCGCTGGCACCTGCACCACCTCAC 4300
 T V L H P K A N A A W H L H H L T
 CCAAAACCAACCCCTCACCCACTTCGTCTACTCCAGCGCCGCCCG 4350
 Q N Q P L T H F V L Y S S A A A
 10 TCCTCGGCAGCCCCGGACAAGGAAACTACGCCGCCAACGCCCTTCCTC 4400
 V L G S P G Q G N Y A A A N A F L
 GACGCCCTCGCCACCCACCGCCACACCCCTCGGCCAACCCGCCACCTCCAT 4450
 D A L A T H R H T L G Q P A T S I
 CGCCTGGGGCATGTGGCACACCACAGCACCCCTACCGGACAACCTGACG 4500
 15 A W G M W H T T S T L T G Q L D
 ACGCCGACCGGGACCGCATCCGCCGGCGGTTCTCCCGATCACGGAC 4550
 D A D R D R I R R G G F L P I T D
 GACGAGGGCATGGGGATGCAT
 D E G

20 The *NheII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 (specific for malonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

25 AGATCTGGCAGCTCGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCCGCCGTGCTCGGCCACGTGGGTGGCGAGGACATCCCCGGACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCCAGCTGCGCAACG 150
 30 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGCACCGGTGTGCGGCTGAACGCCACGGCGGTCTTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGGAAGCTCGGCACGAACGACCGG 250
 F P T P H V L A G K L G D E L T G
 35 CACCCGCGCGCCCGTGTGCCCCGGACCGCGGCCACGGCGGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGCTGCCGGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGACCGACGCCAT 400
 40 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGGCTGGACGTCGACGCCATCACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACCGCGATCGGCAAGACCTTCGTCCGGCACGGTGGCTTCCTC 500
 P D P D A I G K T F V R H G G F L
 45 ACCGGCGCGACAGGCTTCGACGCCGGTCTCGGATCACGCCGCCGA 550
 T G A T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCCGACGCCGGTGTCCCTGGAGACGTCGTGGG 600
 A L A M D P Q Q R V L L E T S W
 AGGCCTTCGAAAGGCCGGCATACCCCGGACTCGACCCGCCAGCGAC 650
 50 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCCTTCTCCACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGCGACCGGCTCGCAGACCAAGTGTGCTCTCCGGC 750
 T D G F G A T G S Q T S V L S G
 55 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTCGTCGCTGGTGGCGCTGCACCGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTCGCCCTGGTCGGCGCGTCACGGTGTGGCGT 900
 60 S G E C S L A L V G G V T V M A

CTCCCGGCGGCTTCGTGGAGTTCTCCGGCAGCGCGGCCCGAC 950
 S P G G F V E F S R Q R G L A P D
 GGCGGGCGAAGGCCTCGCGCGGGTGCAGGACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 5 GGGTGCCGGTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCAACCGCTCTGGCGCGTCCGTGGTCCGGCGTCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGCTGTCGGCGCCGAACGGGCCGTCGAGGAGCAGGGTGT 1150
 10 A S N G L S A P N G P S Q E R V I
 CGGGCAGGCCCTGGCAAACGCCGGCTCACCCGGCGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCCGACCAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 15 GCGGTACTGGCCACCTACGGACAGGAGCGCCACCCCCCTGCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGTCGGCGTCGCCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGAGCTGCCGCCAG 1400
 20 G I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGAGCCGTCGCCGACGTCGACTGGACGCCGGCGCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTGGCCGGCCGGCTGGCCCGAGACCGACCGGCCACGGC 1500
 E L L T S A R P W P E T D R P R
 25 GTGCCGCCGTCTCCTCGTGGGGTGAGCGGCCACCAACGCCACGTCACTC 1550
 R A A V S S F G V S G T N A H V I
 CTGGAGGCCGGACCGGTAACGGAGACGCCCGGGCATGCCCTCCGGTGA 1600
 L E A C P V T E T P A A S P S G D
 CCTTCCCCTGCTGGTGTGGCACGCTCACCGGAAGCGCTCGACGAGCAGA 1650
 30 L P L L V S A R S P E A L D E Q
 TCCGCCGACTGCGCCCTACCTGGACACCACCCCGAGCTCGACCGGGTG 1700
 I R R L R A Y L D T T P D V D R V
 GCCGTGGCACAGACGCTGGCCGGCGCACACACTTCGCCACCAGCGCGCGT 1750
 A V A Q T L A R R T H F A H R A V
 35 GCTGCTCGGTGACACCGTCATCACCAACACCCCCCGCGGACCGGCCAGC 1800
 L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCTACTCCGGCCAGGGCACCCAGCATCCCGCGATGGGC 1850
 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCCGCCGCGTCCCGCTTCGGCGATCCATCAGCAGGT 1900
 40 E Q L A A A F P V F A R I H Q Q V
 GTGGGACCTGCTCGATGTGCCGATCTGGAGGTGAACGAGACCGGTACG 1950
 W D L L D V P D L E V N E T G Y
 CCCAGCCGCCCTGTCGCAATGCAGGTGGCTCTGTCGGGCTGCTGGAA 2000
 A Q P A L F A M Q V A L F G L L E
 45 TCGTGGGGTGTACGACGGACGCCGGTATCGGCCATTGGTGGGTGAGCT 2050
 S W G V R P D A V I G H S V G E L
 TGCGGCTCGTATGTGTCCGGGTGTGGTCGGTGGAGATGCCCTGCACTT 2100
 A A A Y V S G V W S L E D A C T
 TGGTGTGGCGCGGGCTCGTCTGATGCAGGTCTGCCCGCGGGTGGGTG 2150
 50 L V S A R A R L M Q A L P A G G V
 ATGGTCGCTGTCGGTCTGGAGGGATGAGGCCGGCGTGTGGTGA 2200
 M V A V P V S E D E A R A V L G E
 GGGTGTGGAGATGCCGCCGGTCAACGGCCCGTCGTCGGTGGTTCTCTCCG 2250
 G V E I A A V N G P S S V V L S
 55 GTGATGAGGCCGCCGTGTCAGGCCGCCGGAGGGCTGGGAAGTGGACG 2300
 G D E A A V L Q A A E G L G K W T
 CGGCTGGCGACCAGCCACGCCGTTCCATTCCGCCGTATGGAACCCATGCT 2350
 R L A T S H A F H S A R M E P M L
 GGAGGAGTTCCGGCGGTGCCGAAGGCCCTGACCTACCCGACGCCAGG 2400
 60 E E F R A V A E G L T Y R T P Q
 TCTCCATGGCCGTTGGTGTACAGGTGACCACCGCTGAGTACTGGGTGCGG 2450
 V S M A V G D Q V T T A E Y W V R
 CAGGTCCGGACACGGTCCGGTCCGGAGCAGGTGGCCTCGTACCGAGGA 2500
 Q V R D T V R F G E Q V A S Y E D

CGCCGTGTCGAGCTGGGTGCCGACCGGTCACTGGCCCGCCTGGTCG 2550
 A V F V E L G A D R S L A R L V
 ACGGTGTGCGATGCTGCACGGCGACACGAAATCCAGGCCGATCGGC 2600
 D G V A M L H G D H E I Q A A I G
 5 GCCCTGGCCCACCTGTATGTCAACGGCGTCACGGTCGACTGGCCCGCCT 2650
 A L A H L Y V N G V T V D W P A L
 CCTGGGCATGCTCCGGCAACACGGGTGCTGGACCTTCCGACATACGCCT 2700
 L G D A P A T R V L D L P T Y A
 TCCAGCACCAGCGCTACTGGCTCGAGTCGGCACGCCGGCATCCGAC 2750
 10 F Q H Q R Y W L E S A R P A A S D
 GCGGGCCACCCCGTCTGGCTCCGGTATGCCCTCGCCGGTCGCCGG 2800
 A G H P V L G S G I A L A G S P G
 CCGGGTGTTACGGGTTCCGTGCCGACCGGTGCGGACCGCGCGGTGTTCG 2850
 R V F T G S V P T G A D R A V F
 15 TCGCCGAGCTGGCGCTGGCCGCCGCGACTGCGCCACGGTC 2900
 V A E L A L A A A D A V D C A T V
 GAGCGGCTCGACATGCCCTCCGTGCCCGCCGGCCATGGCCGGAC 2950
 E R L D I A S V P G R P G H G R T
 GACCGTACAGACCTGGGTGACGAGCCGGGACGACGGCCGGCGCGGT 3000
 20 T V Q T W V D E P A D D D G R R R
 TCACCGTGCACACCCCGCACCGCGACGCCCGTGGACGCTGACGCCGAG 3050
 F T V H T R T G D A P W T L H A E
 GGGGTGCTGCCGCCCCATGGCACGCCCTGCCGATGCCGACGCCGA 3100
 G V L R P H G T A L P D A A D A E
 25 GTGGCCCCCACCGGGCGCGGTGCCCGCGACGGGCTGCCGGGTGTGGC 3150
 W P P P G A V P A D G L P G V W
 GCCGGGGGGACCGGTCTCGCCGAGGCGAGGTGGACGGACCGGACGGT 3200
 R R G D Q V F A E A E V D G P D G
 TTCGTGGTGCACCCCGACCTGCTCGACGCCGTCTCTCCGCGTCGGCGA 3250
 30 F V V H P D L L D A V F S A V G D
 CGGAAGCCGCCAGCCGGGATGGCGCGACCTGACGGTGCACGCCGTGG 3300
 G S R Q P A G W R D L T V H A S
 ACGCCACCGTACTGCGCGCTGCCCTCACCGGCCACCGACGGAGCCATG 3350
 D A T V L R A C L T R R T D G A M
 35 GGATTGCCGCCCTCGACGGCGCCGGCTGCCGTACTCACCGCGGAGGC 3400
 G F A A F D G A G L P V L T A E A
 GGTGACGCTGCCGGAGGTGGCGTCACCGTCCGGCTCCGAGGAGTCGGACG 3450
 V T L R E V A S P S G S E E S D
 GCCTGCACCGGGTGGAGTGGCTCGCGGTGCGGAGGGCTACGACGGT 3500
 40 G L H R L E W L A V A E A V Y D G
 GACCTGCCGAGGGACATGTCCCTGATCACCGCCGCCACCCGACGACCC 3550
 D L P E G H V L I T A A H P D D P
 CGAGGACATACCCACCCGCCAACACCCGCCAACCCGCGTCCGTACCG 3600
 E D I P T R A H T R A T R V L T
 45 CCCTGCAACACCACTCACCACCGACCCACCCCTCATCGTCCACACC 3650
 A L Q H H L T T T D H T L I V H T
 ACCACCGACCCCGCCGGCCACCGTCACCGGCCCTCACCCGACCGCCA 3700
 T T D P A G A T V T G L T R T A Q
 GAACGAACACCCCCACCGCATCGCCTCATCGAAACCGACCCACACCA 3750
 50 N E H P H R I R L I E T D H P H
 CCCCCTCCCCCTGGCCCAACTCGCCACCCCTCGACCAACCCCACTCCGC 3800
 T P L P L A Q L A T L D H P H L R
 CTCACCCACCAACCCCTCCACCAACCCCACTCACCCCTCCACACCAAC 3850
 L T H H T L H H P H L T P L H T T
 55 CACCCCAACCAACCAACCCCCCTCAACCCGAACACGCCATCATCATCA 3900
 T P P T T T P L N P E H A I I I
 CCGGCGGCTCCGGCACCCCTCGCCGGCATCCTCGCCGCCACCTGAACCCAC 3950
 T G G S G T L A G I L A R H L N H
 CCCCACACCTACCTCTCCCGCACCCCAACCCCGACGCCACCCCGG 4000
 60 P H T Y L L S R T P P P D A T P G
 CACCCACCTCCCCCTGCGACGTCGGCGACCCCCACCAACTCGCCACCAACCC 4050
 T H L P C D V G D P H Q L A T T
 TCACCCACATCCCCAACCCCTCACCGCCATCTTCCACACCGCCGCCACC 4100
 L T H I P Q P L T A I F H T A A T

CTCGACGGCATCCTCACGCCCTCACCCCCGACCGCCTACCAACCGT 4150
 L D D G I L H A L T P D R L T T V
 CCTCCACCCCAAAGCCAACGCCCTGGCACCTGCACCAACCTCACCCAAA 4200
 L H P K A N A A W H L H H L T Q
 5 ACCAACCCCTCACCCACTTCGTCTACTCCAGCGCCGCCGTCC 4250
 N Q P L T H F V L Y S S A A A V L
 GGCAGCCCCGGACAAGGAAACTACGCCGCCAACGCCCTCCTCGACGC 4300
 G S P G Q G N Y A A A N A F L D A
 CCTCGCCACCCACCGCCACACCCTCGCCAACCCGCCACCTCCATGCCCT 4350
 10 L A T H R H T L G Q P A T S I A
 GGGGCATGTGGCACACCACAGCACCCCTCACGGACAACTCGACGACGCC 4400
 W G M W H T T S T L T G Q L D D A
 GACCGGGACCGCATCCGCCGGCGGTTCTCCGATCACGGACGACGA 4450
 D R D R I R R G G F L P I T D D E
 15 GGGCATGGGATGCAT
 G

The *NheII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 13 (specific for 20 methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCCGCCGTGCTCGGCCACGTGGTGGCGAGGACATCCCCGCGACGGCGC 100
 25 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCCAGCTGCGCAACG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGGGACCGGGTGTGGCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 30 TTCCCGACCCCGACGTGCTCGCCGGGAAGCTCGGCGACGAACTGACCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCCGCGCCCGCTGTGCCCGGACCGCGGCCACGGCCGGTGCACCG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGGTC 350
 35 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGCTGGACGTCGACCGATCTACGACC 450
 T E F P T D R G W D V D A I Y D
 40 CGGACCCCGACCGGATCGGAAGACCTTCGTCCGGCACGGTGGCTTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACGCCGGTCTCGGCATCACGGCGCGCA 550
 T G A T G F D A A E F G I S P R E
 GGCCCTCGCGATGGACCCCGACAGCGGGTGCTCCCTGGAGACGTGCG 600
 45 A L A M D P Q Q R V L L E T S W
 AGGCCTTCGAAAGCGCCGGCATCACCCCGACTCGACCCCGGGCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCTTCTCCTACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 50 CACCGACGGCTTCGGCGCACCGGCTCGCAGACCAAGTGTGCTCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTCTACGGTCTGGAGGGTCCGGCGGTACGGTCACCGAC 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTCGTCGCTGGCGCTGCACCAAGGCCGGCAGTCGCTGCG 850
 55 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTCGCCCTGGTCGGCGGTACGGTGATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCCCGGCGGCTTCGTTGGAGATTCTCCCGCAGCGCGGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 60 GGCCGGCGAAGGCCTTCGGCGGGTGCGGACGGCACGAGCTTCGCCGA 1000

G R A K A F G A G A D G T S F A E
 GGGTGCCGGTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCACACCCTGGCGGTGTCGGTGGTCCGGCGGTCAACCAGGATGGT 1100
 5 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGGCTGTCCGGCGCCGAACGGGCCGTGCAGGAGCAGGGTGT 1150
 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCCAACGCCGGCTCACCCCGCGGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 10 TCGAGGCCACGGCACCCGGCACAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCCACCTACGGACAGGGAGCGCCACCCCCCTGCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGCGTCCGGCGTCGCCG 1350
 15 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGAGCTGCCGCCACG 1400
 G I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGAGCCGTCGCCGACGTCGACTGGACGGCCGGCGCCGT 1450
 L H A D E P S P H V D W T A G A V
 20 CGAACTGCTGACGTGGCCCGGCCGTGGCCCGAGACCGACCGGCCACGGC 1500
 E L L T S A R P W P E T D R P R
 GTGCCGCCGTCTCTCGTTGGGGTGAGCGGCACCAACGCCACGTCATC 1550
 R A A V S S F G V S G T N A H V I
 CTGGAGGCCGACCGGTAAACGGAGACGCCCGGGCATGCCCTCCGGTGA 1600
 25 L E A G P V T E T P A A S P S G D
 CCTTCCCCCTGCTGGTGTCCGGCACGGTCACCGGAAGCGCTCGACGAGCAGA 1650
 L P L L V S A R S P E A L D E Q
 TCCGCCGACTCGCGCCTACCTGGACACCACCCGGACGTCGACGGGTG 1700
 30 I R R L R A Y L D T T P D V D R V
 GCCGTGGCACAGACGCTGGCCCGGGCACACACTTCGCCCACCCGGCGGT 1750
 A V A Q T L A R R T H F A H R A V
 GCTGCTCGGTGACACCGTCATCACCAACACCCCGGGACCGGCCGACG 1800
 L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCTACTCCGGCAGGGCACCCAGCATCCCGCATGGG 1850
 35 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCCGATTCGTCGGTGGTTCGGCGAGCGGATGGCCAGTG 1900
 E Q L A D S S V V F A E R M A E C
 TGCGCGGCCGTTGCGCGAGTTCTGGACTGGGATCTGTTCACGGTTCTGG 1950
 A A A L R E F V D W D L F T V L
 40 ATGATCCGGCGGTGGTGGACCGGGTTGATGTTGTCAGGCCGCTTCTGG 2000
 D D P A V V D R V D V V Q P A S W
 GCGATGATGGTTCCCTGGCCGCGGTGTCAGGCCGCGGCGGTGCGGCC 2050
 A M M V S L A A V W Q A A G V R P
 GGATGCGGTGATCGGCCATTGCAAGGGTGAGATGCCGCAGCTTGTGTGG 2100
 45 D A V I G H S Q G E I A A A C V
 CGGGTGCCTGTCACTACCGCATGCCGCCGGATCGTGCACCTTGCAGC 2150
 A G A V S L R D A A R I V T L R S
 CAGGCGATGCCCGGGGCTGGCGGGCGGGCGCATGGCATCCGTCGC 2200
 Q A I A R G L A G R G A M A S V A
 50 CCTGCCCGCAGGATGTCGAGCTGGTCGACGGGCGCTGGATGCCGCC 2250
 L P A Q D V E L V D G A W I A A
 ACAACGGGCCGCTCCACCGTGATCGCGGGCACCCCGGAAGCGGTGAC 2300
 H N G P A S T V I A G T P E A V D
 CATGTCCTCACCGCTCATGAGGCACAAGGGTGCGGGTGCAGGCCG 2350
 55 H V L T A H E A Q G V R V R R I T
 CGTCGACTATGCCCGCACACCCCGCACGTCGAGCTGATCCCGAAC 2400
 V D Y A S H T P H V E L I R D E
 TACTCGACATCACTAGCGACAGCAGCTCGCAGACCCCGCTGCGCCGTGG 2450
 60 L L D I T S D S S S Q T P L V P W
 CTGTCGACCGTGGACGGCACCTGGGTGACAGGCCGCTGGACGGGAGTA 2500
 L S T V D G T W V D S P L D G E Y
 CTGGTACCGGAACCTGCGTGAACGGTGGTCCACCCCGCGTCAGCC 2550
 W Y R N L R E P V G F H P A V S
 AGTTGCAGGCCAGGGCACCGTGAGGTCAGCGCCAGCCCG 2600

Q L Q A Q G D T V F V E V S A S P
 GTGTTGTCAGGCATGGACGATGTCGTACGGTGCACGCCAGCGC 2650
 V L L Q A M D D D V V T V A T L R
 TCGTACGACGGCGACGCCACCCGGATGCTACCGCCCTGGCACAGGC 2700
 5 R D D G D A T R M L T A L A Q A
 ATGTCACGGCGTCACCGTCGACTGGCCGCCATCCTCGGACCACCA 2750
 Y V H G V T V D W P A I L G T T T
 ACCCGGGTACTGGACCTTCCGACCTACGCCTTCCAACACCAGCGG 2800
 T R V L D L P T Y A F Q H Q R Y W
 10 GCTCGAGTCGGCACGCCCGGCCATCCGACGCCGGCCACCCGTGCTGG 2850
 L E S A R P A A S D A G H P V L
 GCTCCGGTATGCCCTGCCGGTGCACGCCGGGGTGTTCACGGGTTCC 2900
 G S G I A L A G S P G R V F T G S
 GTGCCGACCGGGTGCACGCCGGTGTTCGCGAGCTGGCGCTGGC 2950
 15 V P T G A D R A V F V A E L A L A
 CGCCGCGGACGCCGTGACTGCGCACGGTCGAGCGGCTCGACATCGC 3000
 A A D A V D C A T V E R L D I A
 CCGTGCCCCGGCCGGCCGGCATGGCCGGACGACCGTACAGACCTGGGTC 3050
 20 S V P G R P G H G R T T V Q T W V
 GACGAGCCGGGGACGACGGCCGGCGCGGTTCACCGTGCACACCCGCAC 3100
 D E P A D D G R R R F T V H T R T
 CGGCCACGCCCGTGGACGCTGCACGCCGAGGGGGTGCTGCACCCCATG 3150
 G D A P W T L H A E G V L R P H
 GCACGGCCCTGCCGATGCGGCCACGCCGAGTGGCCCCACCGGGCG 3200
 25 G T A L P D A A D A E W P P P G A
 GTGCCCGCGACGGGCTGCCGGGTGTGTGGCGCCGGGGACAGGTCTT 3250
 V P A D G L P G V W R R G D Q V F
 CGCCGAGGCCGAGGTGGACGGACCGACGGTTCGTGGTGCACCCGACC 3300
 A E A E V D G P D G F V V H P D
 30 TGCTCGACGCCGTCTTCTCCGCGTCGGCGACGGAAGCCGCCAGCCGGC 3350
 L L D A V F S A V G D G S R Q P A
 GGATGGCGCGACCTGACGGTGCACGCCGCGACGCCACCGTACTGCCGC 3400
 G W R D L T V H A S D A T V L R A
 CTGCCCTCACCGCGCACCGACGGAGCCATGGGATTGCCGCTTCGACG 3450
 35 C L T R R T D G A M G F A A F D
 GCGCCGGCCTGCCGGTACTCACCGCGAGGGCGGTGACGCTGCCGGAGGTG 3500
 G A G L P V L T A E A V T L R E V
 GCGTCACCGTCCGGCTCCGAGGAGTCGGACGCCCTGCACCGGTTGGAGTG 3550
 A S P S G S E E S D G L H R L E W
 40 GCTCGCGGTGCCGAGGCCGTACGACGGTACCTGCCGAGGGACATG 3600
 L A V A E A V Y D G D L P E G H
 TCCTGATCACCGCCGCCACCCGACGACCCCGAGGACATACCCACCGC 3650
 V L I T A A H P D D P E D I P T R
 GCCCACACCCCGCCACCCCGTCTGACGCCCTGCAACACCCACCTCAC 3700
 45 A H T R A T R V L T A L Q H H L T
 CACCAACCGACCAACCCCTCATCGTCACACCACCGACCCCGCCGGCG 3750
 T T D H T L I V H T T D P A G
 CCACCGTCACCGGCCCTCACCCGACCGCCAGAACGAACACCCACCGC 3800
 A T V T G L T R T A Q N E H P H R
 50 ATCCGCCCTCATCGAAACCGACCCACCCCCCTCCCCCTGGCCCA 3850
 I R L I E T D H P H T P L P L A Q
 ACTCGCCACCCCTCGACCAACCCACCTCCGCCCTCACCCACCCCTCC 3900
 L A T L D H P H L R L T H H T L
 ACCACCCCCACCTCACCCCCCTCCACACCACCCACCCACCCACCC 3950
 55 H H P H L T P L H T T T P P T T T
 CCCCTCAACCCCGAACACGCCATCATCACCGGGGGCTCCGGCACCC 4000
 P L N P E H A I I I T G G S G T L
 CGCCGGCATCTCGCCGCCACCTGAACCAACCCCCACACCTACCTCT 4050
 A G I L A R H L N H P H T Y L L
 60 CCCGCACCCACCCCCCGACGCCACCCCGGACCCACCTCCCTGGC 4100
 S R T P P P D A T P G T H L P C D
 GTCGGCGACCCCCACCAACTCGCCACCCCTCACCCACATCCCCAAC 4150
 V G D P H Q L A T T L T H I P Q P
 CCTCACCGCCATCTTCCACACCGCCGCCACCCCTCGACGACGGCATCCTCC 4200

```

L T A I F H T A A T L D D G I L
ACGCCCTCACCCCCGAECCGCTCACCAACCGTCTCCACCCCAAAGCCAAC 4250
H A L T P D R L T T V L H P K A N
GCCGCCTGGCACCTGCACCACCTCACCCAAAACCAACCCCTCACCCACTT 4300
5   A A W H L H L T Q N Q P L T H F
CGTCCTCTACTCCAGCGCCGCCGCCGTCTCGCAGCCCCGGACAAGGAA 4350
V L Y S S A A A V L G S P G Q G
ACTACGCCGCCAACGCCCTCCTCGACGCCCTGCCACCCACCGCCAC 4400
N Y A A A N A F L D A L A T H R H
10  ACCCTCGGCCAACCGCCACCTCCATCGCCTGGGCATGTGGCACACCAC 4450
T , G Q P A T S I A W G M W H T T
CAGCACCCCTACCGGACAACCTCGACGACGCCGACCGGGACCGCATCCGCC 4500
S T L T G Q L D D A D R D R I R
GCGCGGTTTCTCCGATCACGGACGACGAGGGCATGGGGATGCAT
15  R G G F L P I T D D E G

```

Phage KC515 DNA was prepared using the procedure described in Genetic Manipulation of *Streptomyces*, A Laboratory Manual, edited by D. Hopwood *et al.* A phage suspension prepared from 10 plates (100 mm) of confluent plaques of KC515 on 20 *S. lividans* TK24 generally gave about 3 µg of phage DNA. The DNA was ligated to circularize at the cos site, subsequently digested with restriction enzymes *Bam*HI and *Pst*I, and dephosphorylated with SAP.

Each module 8 cassette described above was excised with restriction enzymes *Bgl*II and *Nsi*I and ligated into the compatible *Bam*HI and *Pst*I sites of KC515 phage 25 DNA prepared as described above. The ligation mixture containing KC515 and various cassettes was transfected into protoplasts of *Streptomyces lividans* TK24 using the procedure described in Genetic Manipulation of *Streptomyces*, A Laboratory Manual edited by D. Hopwood *et al.* and overlaid with TK24 spores. After 16-24 hr, the plaques were restreaked on plates overlaid with TK24 spores. Single plaques were picked and 30 resuspended in 200 µL of nutrient broth. Phage DNA was prepared by the boiling method (Hopwood *et al.*, *supra*). The PCR with primers spanning the left and right boundaries of the recombinant phage was used to verify the correct phage had been isolated. In most cases, at least 80% of the plaques contained the expected insert. To 35 confirm the presence of the resistance marker (thiostrepton), a spot test is used, as described in Lomovskaya *et al.* (1997), in which a plate with spots of phage is overlaid with mixture of spores of TK24 and phiC31 TK24 lysogen. After overnight incubation, the plate is overlaid with antibiotic in soft agar. A working stock is made of all phage containing desired constructs.

Streptomyces hygroscopicus ATCC 14891 (see US Patent No. 3,244,592, issued 40 5 Apr 1966, incorporated herein by reference) mycelia were infected with the recombinant phage by mixing the spores and phage (1×10^8 of each), and incubating on R2YE agar (Genetic Manipulation of *Streptomyces*, A Laboratory Manual, edited by D.

Hopwood *et al.*) at 30°C for 10 days. Recombinant clones were selected and plated on minimal medium containing thiostrepton (50 µg/ml) to select for the thiostrepton resistance-conferring gene. Primary thiostrepton resistant clones were isolated and purified through a second round of single colony isolation, as necessary. To obtain 5 thiostrepton-sensitive revertants that underwent a second recombination event to evict the phage genome, primary recombinants were propagated in liquid media for two to three days in the absence of thiostrepton and then spread on agar medium without thiostrepton to obtain spores. Spores were plated to obtain about 50 colonies per plate, and thiostrepton sensitive colonies were identified by replica plating onto thiostrepton 10 containing agar medium. The PCR was used to determine which of the thiostrepton sensitive colonies reverted to the wild type (reversal of the initial integration event), and which contain the desired AT swap at module 8 in the ATCC 14891-derived cells. The PCR primers used amplified either the KS/AT junction or the AT/DH junction of the wild-type and the desired recombinant strains. Fermentation of the recombinant strains, 15 followed by isolation of the metabolites and analysis by LCMS, and NMR is used to characterize the novel polyketide compounds.

Example 2

Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-506

20 The present invention also provides the 13-desmethoxy derivatives of FK-506 and the novel PKS enzymes that produce them. A variety of *Streptomyces* strains that produce FK-506 are known in the art, including *S. tsukubaensis* No. 9993 (FERM BP-927), described in U.S. Patent No. 5,624,852, incorporated herein by reference; *S. hygroscopicus* subsp. *yakushimaensis* No. 7238, described in U.S. patent No. 4,894,366, 25 incorporated herein by reference; *S. sp.* MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference; and *S. sp.* MA 6548, described in Motamedi *et al.*, 1998, "The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506," *Eur. J. Biochem.* 256: 528-534, and Motamedi *et al.*, 1997, "Structural organization of a multifunctional polyketide synthase involved in the 30 biosynthesis of the macrolide immunosuppressant FK-506," *Eur. J. Biochem.* 244: 74-80, each of which is incorporated herein by reference.

The complete sequence of the FK-506 gene cluster from *Streptomyces* sp. MA6548 is known, and the sequences of the corresponding gene clusters from other FK-506-producing organisms is highly homologous thereto. The novel FK-506 recombinant 35 gene clusters of the present invention differ from the naturally occurring gene clusters in

that the AT domain of module 8 of the naturally occurring PKSs is replaced by an AT domain specific for malonyl CoA or methylmalonyl CoA. These AT domain replacements are made at the DNA level, following the methodology described in Example 1.

5 The naturally occurring module 8 sequence for the MA6548 strain is shown below, followed by the illustrative hybrid module 8 sequences for the MA6548 strains.

```

GCATGGCGCTGTACGAGGCAGCAGCGCACCGGAAGTCCCCTGGTGGTG 50
M R L Y E A A R R T G S P V V V
10      GGGGCCGCCCTCGACGACGCCGGACGTGCCGCTGCTGCGCGGCTGCG 100
A A A L D D A P D V P L L R G L R
GCGTACGACCGTCGGCGTCCGCCGCTCCGGGAACGCTCTCGCCGACC 150
R T T V R R A A V R E R S L A D
GCTCGCCGCTGCTGCCCGACGACGCCGGACGCCCTCCCTCGCGTTCG 200
R S P C C P T T S A P T P P S R S
15      TCCTGGAACAGCACCGCCACCGTGGCTCGGCCACCTGGCGCCGAAGACAT 250
S W N S T A T V L G H L G A E D I
CCCGCGACGACGACGTTCAAGGAACCTGGCATCGACTCGCTACCGCGG 300
P A T T T F K E L G I D S L T A
TCCAGCTGCGCAACGCGCTGACCAACGGCACCAGCGTACGCCCTAACGCC 350
20      V Q L R N A L T T A T G V R L N A
ACAGCGGTCTTCGACTTCCGACGCCGCGCGCTGCCGCGAGACTCGG 400
T A V F D F P T P R A L A A R L G
CGACGAGCTGGCCGGTACCCGCGCGCCGTCGCGGCCCGGACCGCGGCA 450
D E L A G T R A P V A A R T A A
25      CCGGCGCCGCCACGACGAACCGCTGGCGATCGTGGGCATGGCCTGCGT 500
T A A A H D E P L A I V G M A C R
CTGCCGGGGCGGGGTCGGCGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
L P G G V A S P Q E L W R L V A S
CGGCACCGACGCCATCACGGAGTCCCCGGGACCGCGGCTGGGACGTGG 600
30      G T D A I T E F P A D R G W D V
ACCGCGCTCTACGACCCGGACCCCGACGCGATCGGCAAGACCTTCGTCGG 650
D A L Y D P D P D A I G K T F V R
CACGGCGGCTTCCTCGACGGTGCACCGGCTTCGACGCCGCGTTCTCGG 700
H G G F L D D G A T G F D A A F F G
35      GATCAGCCCCGCGGAGGCCCTGGCATGGACCCGCAGAACGGGTGCTCC 750
I S P R E A L A M D P Q Q R V L
TGGAGACGTCCTGGGAGGGCTTCGAAAGCGCGGGCATACCCCGACGCG 800
L E T S W E A F E S A G I T P D A
GCGCGGGGAGCGACACCGCGTGTTCATCGGCGCGTCTCCTACGGGTA 850
40      A R G S D T G V F I G A F S Y G Y
CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGCAGACCA 900
G T G A D T N G F G A T G S Q T
GCGTGCCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
S V L S G R L S Y F Y G L E G P S
45      GTCACGGTCGACACCGCCCTGCTCGTCGTCACTGGTCGCCCTGCACCG 1000
V T V D T A C S S S L V A L H Q A
AGGGCAGTCCTCGCCTCGGGCGAATGCTCGCTGCCCTGGTCGGCGGTG 1050
G Q S L R S G E C S L A L V G G
50      TCACGGTGATGGCGTCGCCCGGGATTCTCGTCGAGTTCTCCCGCAGCGC 1100
V T V M A S P G G F V E F S R Q R
GGGCTCGCGCCGGACGGCGGGCGAAGGCGTTCGGCGCGGGCGCGGACGG 1150
G L A P D G R A K A F G A G A D G
TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTCGAGCGGGCTCTCGG 1200
T S F A E G A G A L V V E R L S
55      ACGGAGCGCCACGGCACACCGTCCCTGCCCTCGTACCGGGCTCCCG 1250
D A E R H G H T V L A L V R G S A
GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCTC 1300
A N S D G A S N G L S A P N G P S
CCAGGAACCGCGTCATCCACCAAGGCCCTCGCGAACCGCAAACCTACCCCG 1350

```

Q E R V I H Q A L A N A K L T P
 CCGATGTCACGCCGGTCGAGGCCACGGCACCCGCCTCGGCAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCACGGCGCTGCTCGGCACGTACGGACAGGACC 1450
 5 P I E A Q A L L A T Y G Q D R A T
 GCCCTGCTCGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTCGCCGGGATCATCAAGATGGTCAGGCCATCGGCAC 1550
 A S G V A G I I K M V Q A I R H G
 10 GAACTGCCGCCACACTGCACGCCGAGGCCGCGCACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCTGAGCCTGACGTCGGCCGGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCGCCGCGCTGCCGCTCGCTCGTCTGGCGTGGCCGGAC 1700
 15 T G R P R R A A V S S F G V S G T
 AACGCCACATCATCCTTGAGGCAGGACCGGTCAAACGGGACCGGTCGA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGACGGTCAAGTAGGACCGGTGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 20 GACCGCTCCCCCGCGCCGCGCTCAGCACGGCGAAGACCTTCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCTCCCCGGAGGCACCTGACGAGCAGATCGGGCGCCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGGCCTATCTGACACCGGCCGGCGTCAACGGGCCGTGGCGC 1950
 25 R A Y L D T G P G V D R A A V A
 AGACACTGGCCCGCGTACGCACCTCACCCACCGGGCGTACTGCTCGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATCGCGCTCCCCCGCGACAGGCCGACGAACTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 30 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCATGGCGAGCACTCG 2100
 V Y S G Q G T Q H P A M G E Q L
 CGGCCGCGTCCCCGTGTCGCCATGCCCTGGCACGACGCCGCTCCGACGG 2150
 A A A F P V F A D A W H D A L R R
 CTCGACGACCCGACCCGACGCCGACGCCAACACGGAGCCAGCACCGCTCTT 2200
 35 L D D P D P H D P T R S Q H T L F
 CGCCCACCAAGCGCGCTTCAACGCCCTCCCTGAGGTCTGGACATCACGC 2250
 A H Q A A F T A L L R S W D I T
 CGCACGCCGTACGCCACTCGCTCGCGAGATCACGCCGCGTACGCC 2300
 P H A V I G H S L G E I T A A Y A
 40 GCCGGGATCTGTCGCTGACGACGCCCTGACCTGATCACCAACCGGTGC 2350
 A G I L S L D D A C T L I T T R A
 CCGCCTCATGCACACGCTTCCGCCCGCCATGGTCACCGTGCTGA 2400
 R L M H T L P P P G A M V T V L
 CCAGCGAGGAGGAGGCCGTCAGCGCTGCCGGCGTGGAGATCGCC 2450
 45 T S E E E A R Q A L R P G V E I A
 GCGGTCTCGGCCACTCCGCTCGTCTCGGGCGACGAGGACGCCGT 2500
 A V F G P H S V V L S G D E D A V
 GCTCGACGTCGCACAGCGCTCGGCATCCACCGCTCTGCCCGCCGC 2550
 L D V A Q R L G I H H R L P A P
 50 ACGCGGGCCACTCCGCGCACATGGAACCGTGGCCGAGCTGCTCGCC 2600
 H A G H S A H M E P V A A E L L A
 ACCACTCGCGAGCTCCGTTACGACGGCCCCACACCGCCATCCGAACGA 2650
 T T R E L R Y D R P H T A I P N D
 CCCCACCAACCGCCGAGTACTGGGGCGAGCAGGTCCGCAACCCCGTGT 2700
 55 P T T A E Y W A E Q V R N P V L
 TCCACGCCACACCCAGCGGTACCCGACGCCGTGTCGAGATCGGC 2750
 F H A H T Q R Y P D A V F V E I G
 CCCGGCCAGGACCTCTCACCGCTGGTCAGGGCATGCCCTGCAGAACGG 2800
 P G Q D L S P L V D G I A L Q N G
 60 CACGGCGGACGGTGCACGCCGCTGCCCGCCCTCTCA 2850
 T A D E V H A L H T A L A R L F
 CACGCGGCCACGCTGACTGGTCCCGCATCCTCGGGGTGTCGCGG 2900
 T R G A T L D W S R I L G G A S R
 CACGACCTGACGTCCCTCGTACCGCGTCCAGCGCCGTCCTACTGGAT 2950

H D P D V P S Y A F Q R R P Y W I
 CGAGTCGGCTCCCCGGCCACGGCCGACTCGGGCACCCTCGGCA 3000
 E S A P P A T A D S G H P V L G
 CCGGAGTCGCCGTGCCGGTCCCGGGCCGGGTGTTACGGGTCCC GTG 3050
 5 T G V A V A G S P G R V F T G P V
 CCCGCCGGTCCGGACCCGGCGGTGTTCATCGCCGAACGGCGCTCGCCGC 3100
 P A G A D R A V F I A E L A L A A
 CGCCGACGCCACCGACTCGGCCACGGTCGAACAGCTCGACGTAC CCTCCG 3150
 10 A D A T D C A T V E Q L D V T S
 TGCCCCGGCGGATCCGCCCGGGCAGGGCCACCGCGCAGACCTGGGT CGAT 3200
 V P G G S A R G R A T A Q T W V D
 GAACCCGCCGCCGACGGGCGCCGCTCACCGTCCACACCCGGT CGG 3250
 E P A A D G R R R F T V H T R V G
 CGACGCCCGTGGACGCTGCACGCCGAGGGGGTTCTCCGCCCCGGCCG 3300
 15 D A P W T L H A E G V L R P G R
 TGGCCCCAGCCCGAAGCCGTCGACACCGCCTGGCCCCCGCCGGCGGTG 3350
 V P Q P E A V D T A W P P P G A V
 CCCGC GGACGGGCTGCCCGGGCGTGGCGACGCCGGACAGGTCTCGT 3400
 P A D G L P G A W R R A D Q V F V
 20 CGAACGCCGAAGTCGACAGCCCTGACGGCTTCGTGGCACACCCGAC CTGC 3450
 E A E V D S P D G F V A H P D L
 TCGACGCCGGTCTTCTCCGCCGGTCGGCGACGGGAGGCCAGCCGACCGA 3500
 L D A V F S A V G D G S R Q P T G
 25 TGGCGCGACCTCGCGGTCCACCGCTCGGACGCCACCGTGCTGCGCGC CTG 3550
 W R D L A V H A S D A T V L R A C
 CCTCACCCGCCCGCACAGTGGTGTGCGTGGAGCTGCCGCCTCGACGGT G 3600
 L T R R D S G V V E L A A F D G
 CCGGAATGCCGGTGCTCACCGCGGAGTCGGTGACGCTGGCGAGGGTC GCG 3650
 30 A G M P V L T A E S V T L G E V A
 TCGGCAGGCCGGATCCGACGAGTCGGACGGTCTGCTGGCTTGAGTGGTT 3700
 S A G G S D E S D G L L R L E W L
 GCCGGTGGCGGAGGCCACTACGACGGTGC CGACGAGTGGCCGAGGGCT 3750
 P V A E A H Y D G A D E L P E G
 ACACCCCTCATCACCGCCACACACCCGACGACCCGACGACCCACCAAC 3800
 35 Y T L I T A T H P D D P D D P T N
 CCCCCACAACACACCCACACGCACCCACACACAAACACACCGTCCCTCAC 3850
 P H N T P T R T H T Q T T R V L T
 CGCCCTCCAACACCCACCTCATCACCAACACACCCCTCATCGTCCACA 3900
 40 A L Q H H L I T T N H T L I V H
 CCACCAACGCCACCCAGGC CGCCGCTCACCGGCCTCACCGCACCGA 3950
 T T T D P P G A A V T G L T R T A
 CAAAACGAACACCCGGCGCATCCACCTCATCGAAACCCACACCCCAA 4000
 Q N E H P G R I H L I E T H H P H
 CACCCCACTCCCCCTCACCAACTCACCAACCTCCACCAACCCACCTAC 4050
 45 T P L P L T Q L T T L H Q P H L
 GCCTCACCAACAACACCCCTCCACACCCCCCACCTCACCCCATCACCA C 4100
 R L T N N T L H T P H L T P I T T
 CACCACAAACACCACCAACCCACCCCCAACACCCACCCCTCAACCCCAA 4150
 H H N T T T T P N T P P L N P N
 50 CCACGCCATCCTCATCACCGCGGCTCCGGCACCCCTCGCCGGCATCCTCG 4200
 H A I L I T G G S G T L A G I L
 CCCGCCACCTCAACCACCCCCCACACCTACCTCCTCTCCGCACACCA C 4250
 A R H L N H P H T Y L L S R T P P
 CCCCCCACACACCCGGCACCCACATCCCTCGCACCTCACCGACCCAC 4300
 55 P P T T P G T H I P C D L T D P T
 CCAAATCACCAAGCCCTCACCCACATACCACACACCCCTCACCGCATCT 4350
 Q I T Q A L T H I P Q P L T G I
 TCCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCC C 4400
 F H T A A T L D D A T L T N L T P
 60 CAACACCTCACCAACCCCTCCAACCAAAGCCGACGCCGCTGGCACCT 4450
 Q H L T T T L Q P K A D A A W H L
 CCACCAACACCCAAAACCAACCCCTCACCCACTTCGTCTACTCCA 4500
 H H H T Q N Q P L T H F V L Y S
 GCGCCGCCGCCACCCCTGGCAGCCCCGGCAAGCCAACACTACGCCGCC 4550

S A A A T L G S P G Q A N Y A A A
 AACGCCTTCCTCGACGCCCTGCCACCCACCGCCACACCAAGGACAACC 4600
 N A F L D A L A T H R H T Q G Q P
 CGCCACCACCATCGCCTGGGCATGTGGCACACCACCAACTCACCA 4650
 5 A T T I A W G M W H T T T T L T
 GCCAACTCACCGACAGCGACCGCAGCGCATCCGCCGCGCGCTCCTG 4700
 S Q L T D S D R D R I R R G G F L
 CCGATCTCGGACGACGAGGGCATGC
 P I S D D E G M

10 The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

GCATGGGGCTGTACGAGGGGGCACGGCGCACCGGAAGTCCCCTGGTGGTG 50
 M R L Y E A A R R T G S P V V V
 15 GCGGCCGCGCTCGACGACGCCGGACGTGCCCTGCTGCCGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCTGCCGCCGCTCCGGAAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCCCTCGCGTTG 200
 20 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGCTCGGCCACCTGGGCCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACCTCGGCATCGACTCGCTCACCGCG 300
 P A T T T F K E L G I D S L T A
 25 TCCAGCTGCGAACCGCGCTGACCACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCCGCGCCCTGCCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCCGCGCCCGTCCGGCCCGAACCGCGGCCA 450
 30 D E L A G T R A P V A A R T A A
 CCGCGGCCGCGCACGACGAACCGCTGGCGATCGTGGCATGGCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGGTCGCGTCCGACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 35 CGGCACCGACGCCATCACGGAGTTCCCCCGCGGACCGCGGCTGGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCGACCGATCGGCAAGACCTTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGTTCCCTCGACGGTGCACCGGTTCGACGCCGGTTCTCGG 700
 40 H G G F L D G A T G F D A A F F G
 GATCAGCCCCGCGGAGGCCCTGGCCATGGACCCGACGAAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCCTGGGAGGCCTCGAAAGCGCGGGCATACCCCGGACCG 800
 L E T S W E A F E S A G I T P D A
 45 GCGCGGGGCAGCGACACCGCGTGTTCATCGGCCGCGTCTCCCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAAACGGCTCGGCCGACAGGGTCGCAGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTTCG 950
 50 S V L S G R L S Y F Y G L E G P S
 GTCACGGTCGACACCGCTCGTCGACTGGTCGCCCTGCCACAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTCGCGCTGGCGAATGCTCGCTCGCCCTGGTCCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 55 TCACGGTGTGGCGTCGCCCGGGATTCTCGTCGAGTTCTCCCGGACCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCCGGACGGCGGGCGAAGGGCTCGCCGCGGGCGCGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCAGGGCGCCGGTCCCTGGTGGTCGAGCGGCTCTCCG 1200
 60 T S F A E G A G A L V V E R L S
 ACGCGGAGCAGGCCACGGCCACACCGTCCCTCGCCCTCGTACGCCGCTCCG 1250
 D A E R H G H T V L A L V R G S A

GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACGCGTCATCCACCAGGCCCTCGCGAACCGAAGCTACCCCCCG 1350
 Q E R V I H Q A L A N A K L T P
 5 CCGATGTCGACCGCGTCAGGCCGACGGCACCGGCACCCGCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCGAGGCCGCTGCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCTGCTCGGCTCGTGAAGTCGAACATCGGGCACGCCAGGCC 1500
 10 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTGCCGGGATCATCAAGATGGTCAGGCCATCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCACACTGCACGCCAGGCCGTCGCCGACGTGACTG 1600
 E L P P T L H A D E P S P H V D W
 15 GACGCCGGTGCCGTCGAGCTCTGACGTCGGCCGGCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCCTAGGCCGGCAGGCCGTCGTCCTCGGGATCAGTGGCACC 1700
 T G R P R R A G V S S F G I S G T
 AACGCCACGTCATCTGAAAGGCCACCCCCACTCAGGCTGCCGACAA 1750
 20 N A H V I L E S A P P T Q P A D N
 CGCGGTGATCGAGCGGGCACCGGAGTGGGTGCCGTTGGTATTCGGCCA 1800
 A V I E R A P E W V P L V I S A
 GGACCCAGTCGGCTTGACTGAGCACGAGGGCCGGTGCCTGCGTATCTG 1850
 R T Q S A L T E H E G R L R A Y L
 25 GCGCGTCGCCGGGGTGGATATGCCGCTGTCGACGCTGGCGAT 1900
 A A S P G V D M R A V A S T L A M
 GACACGGTCGGTGGTCGAGCACCGTCCGTCGCTGGGAGATGACACCG 1950
 T R S V F E H R A V L L G D D T
 TCACCGGCACCCCTGTCGACCCCTGGCGGTGTCGCTCTCCCGGA 2000
 30 V T G T A V S D P R A V F V F P G
 CAGGGGTCGACCGTGCTGGCATGGTGAGGAACCTGGCCGCCGTTCCC 2050
 Q G S Q R A G M G E E L A A A F P
 CGTCTCGCGCGGATCCATCAGCAGGTGTGGGACCTGCTCGATGTGCCG 2100
 V F A R I H Q Q V W D L L D V P
 35 ATCTGGAGGTGAACGAGACCGTTACGCCAGCCGGCCCTGTCGCAATG 2150
 D L E V N E T G Y A Q P A L F A M
 CAGGTGGCTCTGTCGGCTGCTGGATCTGGGTGTACGACCGGACGC 2200
 Q V A L F G L L E S W G V R P D A
 GGTGATCGGCCATTGGTGGGTGAGCTTGGCTGCGTATGTGTCCGGGG 2250
 40 V I G H S V G E L A A A Y V S G
 TGTGGTCGTTGGAGGATGCCCTGCACTTGGTGTGGCGCGGGCTGCTG 2300
 V W S L E D A C T L V S A R A R L
 ATGCAGGCTCTGCCGGGGTGGGTGATGGTCGCTGCCGGTCTCGGA 2350
 M Q A L P A G G V M V A V P V S E
 45 GGATGAGGCCGGGGCGCTGGGTGAGGTGTGGAGATGCCGCCGTC 2400
 D E A R A V L G E G V E I A A V
 ACGGCCCGTCGCTGGTGGTCTCTCCGGTGATGAGGCCGGCTGCTGAG 2450
 N G P S S V V L S G D E A V L Q
 GCCCGGGAGGGCTGGGAAGTGGGACGCCGCTGGCACCAGCCACGCC 2500
 50 A A E G L G K W T R L A T S H A F
 CCATTCCGCCGTATGGAAACCCATGCTGGAGGGAGTCCGGGGCGTCGCC 2550
 H S A R M E P M L E E F R A V A
 AAGGCCCTGACCTTACCGGACGCCGCAGGTCTCCATGGCCGGTGGTGTACAG 2600
 E G L T Y R T P Q V S M A V G D Q
 55 GTGACCACCGCTGAGTACTGGGTGGCGAGGTCCGGGACACGGTCCGGT 2650
 V T T A E Y W V R Q V R D T V R F
 CGCGAGCAGGTGGCTCGTACGAGGACGCCGTGGTGTGAGCTGGGTG 2700
 G E Q V A S Y E D A V F V E L G
 CCGACCGGTCACTGGCCCGCTGGTCGACGGTGTGCGATGCTGCCGGC 2750
 60 A D R S L A R L V D G V A M L H G
 GACCACGAAATCCAGGCCCGATCCGGCCCTGGCCCACCTGTATGTCAA 2800
 D H E I Q A A I G A L A H L Y V N
 CGCGTCACGGTCGACTGGCCCGCCTGGCGATGCTCCGGAACACAC 2850
 G V T V D W P A L L G D A P A T

GGGTGCTGGACCTCCGACATACGCCTTCCAGCACCAGCGCTACTGGCTC 2900
 R V L D L P T Y A F Q H Q R Y W L
 GAGTCGGCTCCCCGGCACGGCCACTCGGGCACCCCGTCTCGGCAC 2950
 E S A P P A T A D S G H P V L G T
 5 CGGAGTCGCCGTGCCGGTCGCCGGCCGGGTGTTCACGGTCCCGTGC 3000
 G V A V A G S P G R V F T G P V
 CCGCCGGTGCAGCGCGCGTGTTCATCGCCAACTGGCGCTCGCC 3050
 P A G A D R A V F I A E L A L A A
 GCCGACGCCACCGACTGCCACGGTCGAACAGCTCGACGTCACCTCCGT 3100
 10 A D A T D C A T V E Q L D V T S V
 GCCCGCGGATCCGCCCGGGCAGGGCCACCGCGCAGACCTGGTCGATG 3150
 P G G S A R G R A T A Q T W V D
 AACCCGCCGCGACGGGCGCGCCGTTCACCGTCCACACCCCGCGTCGGC 3200
 E P A A D G R R F T V H T R V G
 15 GACGGCCCGTGGACGCTGCACGCCGAGGGGTCTCCGCCCGCGGT 3250
 D A P W T L H A E G V L R P G R V
 GCCCAGCCCAGGCCGTGACACCCGCTGGCCCCGGCGCGGTGC 3300
 P Q P E A V D T A W P P P G A V
 CCGCGGACGGGCTGCCGGGTGGCGACGCCGAGGGTCTCGTC 3350
 20 P A D G L P G A W R R A D Q V F V
 GAAGCCGAAGTCGACAGCCTGACGGCTTCGGCACACCCGACCTGCT 3400
 E A E V D S P D G F V A H P D L L
 CGACCGGGTCTCTCCGGTGGCGACGGGAGCCGCCAGCCGACCGGAT 3450
 D A V F S A V G D G S R Q P T G
 25 GGCACGCGACCTCGCGGTGCACCGTGGACGCCACCGTGCTCGCGCCTGC 3500
 W R D L A V H A S D A T V L R A C
 CTCACCCGCCGACAGTGGTGTGGAGCTGCCGCCCTCGACGGTGC 3550
 L T R R D S G V V E L A A F D G A
 CGGAATGCCGGTGCTCACCGGGAGTCGGTGACGCTGGCGAGGTGGT 3600
 30 G M P V L T A E S V T L G E V A
 CGGCAGGCCGATCCGACGAGTCGGACGGTCTGGCTTGAGTGGTTG 3650
 S A G G S D E S D G L L R L E W L
 CCGTGGCGAGGCCACTACGACGGTGGCGACGAGCTGCCGAGGGCTA 3700
 P V A E A H Y D G A D E L P E G Y
 35 CACCTCATACCGCCACACACCCGACGACCCGACGACCCACCAACC 3750
 T L I T A T H P D D P D D P T N
 CCCACAACACACCCACACCGCACACACAAACACACCGCTCTCACC 3800
 P H N T P T R T H T Q T T R V L T
 GCCCTCCAACACCACCTCATACCAACCACACCCCTCATCGTCCACAC 3850
 40 A L Q H H L I T T N H T L I V H T
 CACCAACGACCCCCCAGGCAGCCGCTACCGGCCTACCCGACCGCAC 3900
 T T D P P G A A V T G L T R T A
 AAAACGAACACCCGGCCATCCACCTCATCGAAACCCACACCCAC 3950
 Q N E H P G R I H L I E T H H P H
 45 ACCCCACTCCCCCTCACCAACTCACCAACCCCTCCACCAACCCACCTACG 4000
 T P L P L T Q L T T L H Q P H L R
 CCTCACCAACACCCCTCCACACCCCCACCTCACCCCATCACCAACCC 4050
 L T N N T L H T P H L T P I T T
 ACCACAACACCACACAACCCACCCCCAACACCCCCACCCCTCAACCCCAAC 4100
 50 H H N T T T T P N T P P L N P N
 CACGCCATCCTCATCACGGCGCTCCGGCACCCCTGCCGGCATCCTCGC 4150
 H A I L I T G G S G T L A G I L A
 CCGCACCTCAACCAACCCCCACACCTACCTCCTCCCGACACCAACCC 4200
 R H L N H P H T Y L L S R T P P
 55 CCCCCACCAACCCGGCACCCACATCCCTGCGACCTCACCGACCCAC 4250
 P P T T P G T H I P C D L T D P T
 CAAATCACCAAGCCCTCACCCACATACCAACACCCCTCACCGGATCTT 4300
 Q I T Q A L T H I P Q P L T G I F
 CCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCC 4350
 60 H T A A T L D D A T L T N L T P
 AACACCTCACCAACCCCTCCAACCAAAGCCGACGCCGCTGGCACCTC 4400
 Q H L T T T L Q P K A D A A W H L
 CACCAACACCCAAAACCAACCCCTCACCCACTTCGTCCCTACTCCAG 4450
 H H H T Q N Q P L T H F V L Y S S

CGCCGCCGCCACCCCTCGCAGCCCCGGCCAAGCCAACACTACGCCGCCCA 4500
 A A A T L G S P G Q A N Y A A A
 ACGCCTCCCTCGACGCCCTCGCCACCCACCGCACACCCAAGGACAACCC 4550
 N A F L D A L A T H R H T Q G Q P
 5 GCCACCACCATCGCCTGGGCATGTGGCACACCACCCACACTCACCAG 4600
 A T T I A W G M W H T T T L T S
 CCAACTCACCGACAGCGACCGCATCCGCCGCGGGCTCCTGC 4650
 Q L T D S D R D R I R R G G F L
 CGATCTGGACGACGAGGGCATGC
 10 P I S D D E G M

The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 13 of rapamycin is shown below.

GCATCGGGCTGTACGAGGGGGCACGGCGCACCGGAAGTCCC GTGGTGGTG 50
 15 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGGCACGTGCCGTGCTGCCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCGTGCCGCCGTCCGGGAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 20 GCTCGCCGTGCTGCCGACGACGGCGGCCGACGCCCTCCCTCGCGTTCG 200
 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGCCTGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACTCGGCATCGACTCGCTCACCGCGG 300
 25 P A T T T F K E L G I D S L T A
 TCCAGCTGCGAACCGCGCTGACCAACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCGCGCTGCCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 30 CGACGAGCTGCCGGTACCCGCGGCCGTCGCCGGCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 CCGCGGCCGCGCACGACGAACCGCTGGCGATCGTGGCATGGCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCGGGCGGGGTCGCGTCGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 35 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTTCCCCGGACCGCGGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACCGCATCGCAAGACCTTCGTCGG 650
 D A L Y D P D P D A I G K T F V R
 40 CACGGCGGCTTCCTCGACGGTGCACCGGCTTCGACGCCGCTTCGG 700
 H G G F L D G A T G F D A A F F G
 GATCAGCCCGCGCGAGGCCCTGGCATGGACCCCGCAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGCTCTGGGAGGGCTTCGAAAGCGCGGGCATACCCCGACCG 800
 45 L E T S W E A F E S A G I T P D A
 GCGCGGGGCAGCGACACCCGGCGTGTTCATCGGCCGCTCCTACGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGGGATACCAACGGCTTCGGCGACAGGGTCGCAGACCA 900
 G T G A D T N G F G A T G S Q T
 50 GCGTGTCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 GTCACGGTCGACACCGCCTGCTCGTCACTGGTCGCCCTGCACCAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTCGCCTCGGGCGAATGCTCGCTGCCCTGGTCGGCGGTG 1050
 55 G Q S L R S G E C S L A L V G G
 TCACGGTGATGGCGCTGCCCGGGATTCGTCGAGTTCTCCCGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGCGAAGGCCTTCGGCGCGGGACGG 1150
 G L A P D G R A K A F G A G A D G
 60 TACGAGCTCGCCGAGGGCGCCGGTGCCTGGTGGTCAGCGGGCTCCG 1200
 T S F A E G A G A L V V E R L S
 ACGCGGAGGCCACGGCACACCGTCCCTCGTACCGGGCTCCGCG 1250

D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAGGCCCTCGCGAACCGAAGACTCACCCCCG 1350
 5 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACCGCGTCAGGCGCACGGCACCGGCACCGGCCCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGGCGCAGGCGCTGCGACGTACGGACAGGACCGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 10 GCCCCTGCTGCTCGGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCCG 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTGCCGGGATCATCAAGATGGTCAGGCCATCCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGGACGAGCCGTCGCCGACGTCGACTG 1600
 15 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCGTCAGCTCTGACGTCGGCCGGCGTGCCGGGG 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCCTAGGCGGGCGTGTGTCGTCCTCGGAGTCAGCGGCACC 1700
 T G R P R R A G V S S F G V S G T
 20 AACGCCACGTACCTGGAGAGCGCACCCCCCGCTCAGCCCAGGAGGA 1750
 N A H V I L E S A P P A Q P A E E
 GGCAGCAGCCTGTTGAGACGCCGGTGGTGGCCTCGGATGTGCTGCCGCTGG 1800
 A Q P V E T P V V A S D V L P L
 TGATATCGGCCAACGACCCAGCCGCCCTGACCGAACACGAAGACCGGCTG 1850
 25 V I S A K T Q P A L T E H E D R L
 CGCGCCTACCTGGCGCGTCGCCGGGCGGATATACTGGCTGTGGCATC 1900
 R A Y L A A S P G A D I R A V A S
 GACGCTGGCGGTGACACGGTCGGTGGTGGACACCGCGCGTACTCCTG 1950
 T L A V T R S V F E H R A V L L
 30 GAGATGACACCGTCACCGCACCGCGGTGACCGACCCCAGGATCGTGT 2000
 G D D T V T G T A V T D P R I V F
 GTCTTCCCGGCGAGGGGTGGCAGTGGCTGGGATGGCAGTCACGCG 2050
 V F P G Q G W Q W L G M G S A L R
 CGATTCTGGCGGTGTCACGGTCGGTGGCGAGCGGATGGCGAGTGTGCGGGCGT 2100
 35 D S S V V F A E R M A E C A A A
 TGCAGGCTGGACTGGGATCTGTCACGGTCTGGATGATCCGGCG 2150
 L R E F V D W D L F T V L D D P A
 GTGGTGGACCGGGTTGATGTGGTCCAGCCCGCTCTGGCGATGATGGT 2200
 V V D R V D V V Q P A S W A M M V
 40 TTCCCTGGCCCGGGTGTGGCAGGCAGGCCGGTGTGCGGCCGGATGCGGTGA 2250
 S L A A V W Q A A G V R P D A V
 TCGGCCATTGCGCAGGGTGAGATGCCGCAGCTTGTGTGGCGGGTGGCGT 2300
 I G H S Q G E I A A A C V A G A V
 TCACTACGCGATGCCGCCGGATCGACCTTGCAGCCAGGCGATCGC 2350
 45 S L R D A A R I V T L R S Q A I A
 CGGGGCGCTGGCGGGCCGGCGCATGGCATCCGTCGCCCTGCCCGCGC 2400
 R G L A G R G A M A S V A L P A
 AGGATGTCGAGCTGGTCGACGGGGCTGGATGCCGCCAACACGGGCC 2450
 Q D V E L V D G A W I A A H N G P
 50 GCCTCCACCGTGATGCCGGCACCCCGGAAGCGGTGACCATGTCCTCAC 2500
 A S T V I A G T P E A V D H V L T
 CGCTCATGAGGCACAAGGGGTGGCGGGGATCACCGTCGACTATG 2550
 A H E A Q G V R V R I T V D Y
 CCTCGCACACCCCGCACGTCGAGCTGATCCGCGACGAACACTCGACATC 2600
 55 A S H T P H V E L I R D E L L D I
 ACTAGCGACAGCAGCTCGCAGACCCCGCTCGTGGCGTGGCGACCGT 2650
 T S D S S S Q T P L V P W L S T V
 GGACGGCACCTGGGTGACAGCCCGCTGGACGGGAGTACTGGTACCGA 2700
 D G T W V D S P L D G E Y W Y R
 60 ACCTCGGTGAACCGGGTGGTTCCACCCGCCGTCAAGCCAGTTGCAGGCC 2750
 N L R E P V G F H P A V S Q L Q A
 CAGGGCGACACCGTGGTCGAGGTGACGCCAGCCGGTGTGCA 2800
 Q G D T V F V E V S A S P V L L Q
 GGCAGTGGACGACGATGTCGTACGGTTGCCACCGCTGCGTCGTGACGACG 2850

A M D D D V V T V A T L R R D D
 GCGACGCCACCCGGATGCTCACCGCCCTGGCACAGGCCTATGTCCACGGC 2900
 G D A T R M L T A L A Q A Y V H G
 GTCACCGTCGACTGGCCCGCCATCCTCGGCACCACCAACCCGGGTACT 2950
 5 V T V D W P A I L G T T T T R V L
 GGACCTTCCGACCTACGCCCTCCAACACCAGCGGTACTGGCTCGAGTCGG 3000
 D L P T Y A F Q H Q R Y W L E S
 CTCCCCCGGCCACGGCCGACTCGGGCCACCCCGTCTCGGCACCGGAGTC 3050
 10 A P P A T A D S G H P V L G T G V
 GCCGTGCGCCGGTGTGCGCCGGCCGGGTGTTCACGGTCCCGTGCCCGCCGG 3100
 A V A G S P G R V F T G P V P A G
 TGCGGACCGCGCGGTGTTCATGCCGAACCTGGCGCTGCCGCCGCGACG 3150
 A D R A V F I A E L A L A A A D
 CCACCGACTGCCAACGGTCGAACAGCTCGACGTACCTCCGTGCCCGC 3200
 15 A T D C A T V E Q L D V T S V P G
 GGATCCGCCCGCGCAGGGCCACCGCGCAGACCTGGTCGATGAACCCGC 3250
 G S A R G R A T A Q T W V D E P A
 CGCCGACGGGCGGCCGCTCACCGTCCACACCCCGCGTCGGCGACGCC 3300
 A D G R R R F T V H T R V G D A
 20 CGTGGACGCTGACGCCGAGGGGTTCTCCGCCCGCGCGTGCCCG 3350
 P W T L H A E G V L R P G R V P Q
 CCCGAAGCCGTCGACACCGCCTGGCCCCCGCCGGCGCGTGCCCGCGA 3400
 P E A V D T A W P P P G A V P A D
 CGGGCTGCCGGGCGTGGCGACGCCGGACCAAGGTCTCGTCGAAGCCG 3450
 25 G L P G A W R R A D Q V F V E A
 AAAGTCGACAGCCCTGACGGCTTCGGCACACCCGACCTGCTCGACGCC 3500
 E V D S P D G F V A H P D L L D A
 GTCTCTCCCGGGTCGGCGACGGGAGCCGACCCGACCGGATGGCCCGA 3550
 V F S A V G D G S R Q P T G W R D
 30 CCTCGCGGTGCAACCGTCCGACGCCACCGTGTGCGCGCTGCCCTACCC 3600
 L A V H A S D A T V L R A C L T
 GCCGCGACAGTGGTGTGGAGCTGCCGCCCTCGACGGTGCCGGAATG 3650
 R R D S G V V E L A A F D G A G M
 CCGGTGCTCACCGCGGAGTCGGTGACGCTGGCGAGGTCCGTCGGCAGG 3700
 35 P V L T A E S V T L G E V A S A G
 CGGATCCGACGAGTCGGACGGTCTGGCTTGAGTGGTTGCCGGTGG 3750
 G S D E S D G L L R L E W L P V
 CGGAGGCCCACTACGACGGTGGCGACGAGCTGCCGAGGGCTACACCTC 3800
 A E A H Y D G A D E L P E G Y T L
 40 ATCACCGCCACACACCCGACGACCCGACGACCCACAAACCCCCACAA 3850
 I T A T H P D D P D D P T N P H N
 CACACCCACACGCACCCACACAAACACACCGTCTCACGCCCTCC 3900
 T P T R T H T Q T T R V L T A L
 AACACCACCTCATCACCAACCACACCCCTCATCGTCCACACCACCC 3950
 45 Q H H L I T T N H T L I V H T T T
 GACCCCCCAGGCAGCCGCGCTCACCGCCTCACCGCACCGCACAAAGGA 4000
 D P P G A A V T G L T R T A Q N E
 ACACCCCGGCCATCCACCTCATCGAAACCCACCAACCCCCACACCCAC 4050
 H P G R I H L I E T H H P H T P
 50 TCCCCCTCACCAACTCACCAACCCCTCCACCAACCCACCTACGCCCTCACC 4100
 L P L T Q L T T L H Q P H L R L T
 AACAAACACCCCTCCACACCCCCCACCTCACCCACACCCACCAA 4150
 N N T L H T P H L T P I T T H H N
 CACCAACCACACCCACACCCCCAACACCCCCACCCCTCACCCACACGCCA 4200
 55 T T T T P N T P P L N P N H A
 TCCTCATACCGGGCGGCTCCGGCACCCCTCGCCGGCATCCTCGCCGCCAC 4250
 I L I T G G S G T L A G I L A R H
 CTCAACCACCCCCACACCTACCTCCCTCCGCACACCACCCACCA 4300
 L N H P H T Y L L S R T P P P P T
 60 CACACCCGGCACCCACATCCCTCGGACCTCACCGACCCACCCAAATCA 4350
 T P G T H I P C D L T D P T Q I
 CCCAAGCCCTCACCCACATACCAACACCCCCCTCACCGGATCTTCCACACC 4400
 T Q A L T H I P Q P L T G I F H T
 GCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCCCCAACACCT 4450

A A T L D D A T L T N L T P Q H L
 CACCAACCACCCCTCCAACCAAAGCCGACGCCGCCTGGCACCTCCACCAACC 4500
 T T T L Q P K A D A A W H L H H
 ACACCCAAAACCAACCCCTCACCCACTCGTCTACTCCAGCGCCGCC 4550
 5 H T Q N Q P L T H F V L Y S S A A
 GCCACCCCTCGGCAGCCCCGGCAAGCCAACACTACGCCGCCAACGCC 4600
 A T L G S P G Q A N Y A A A A N A F
 CCTCGACGCCCTCGCCACCCACCGCCACACCCAAGGACAACCGCCACCA 4600
 L D A L A T H R H T Q G Q P A T
 10 CCATCGCCTGGGCATGTGGCACACCAACCAACTCACCAGCCAACTC 4700
 T I A W G M W H T T T T L T S Q L
 ACCGACAGCGACCGCGACCGCATCCGCCGCCGGCTTCCTGCCGATCTC 4750
 T D S D R D R I R R G G F L P I S
 GGACGACGAGGGCATGC
 15 D D E G M

The *NheI-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

20 GCATGCGGCTGTACGAGGCCGACGGCGCACCGGAAGTCCC GTGGTGGTG 50
 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGCCGGACGTGCCGTGCTGCCGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCTGCCGGTCCGGAAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 25 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCCCTCGCGTTG 200
 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGCTCGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACCTCGGACTCGCTCACCGCGG 300
 30 P A T T T F K E L G I D S L T A
 TCCAGCTGCGAACCGCGCTGACCACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L P N A L T T A T G V R L N A
 ACAGCGGTCTCGACTTCCGACGCCGCGCGCTCGCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 35 CGACGAGCTGGCCGGTACCCGCGGCCGTGCGGGCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 CCGCGCCGCGACGACGAAACCGCTGGCGATCGTGGCATGGCGTCCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGGTGCGCTGCCACAGGAGCTGTGGCGTCTCGCGTC 550
 40 L P G G V A A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTTCCCCCGCGGACCGCGGCTGGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACCGATCGCAAGACCTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 45 CACGGCGCTCCCTCGACGGTGCACCGGCTTCGACGCCGGTTCTCGG 700
 H G G F L D G A T G F D A A F F G
 GATCAGCCCGCGAGGCCATGGACCCGAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCCTGGGAGGGCTTCGAAAGCGCGGGCATACCCCGGACCG 800
 50 L E T S W E A F E S A G I T P D A
 GCGCGGGCGACGCCGGCTGGCATCGTACTCGTCCCTACGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGCAGACCA 900
 G T G A D T N G F G A T G S Q T
 55 GCGTGCCTCCGGCCCTCTCGTACTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 GTCACGGTCGACACCGCCCTGCTCGTCACTGGTCGCCCTGCACCAAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTCGCGCTGGCGAATGCTCGCTCGCCCTGGTCGGCGGTG 1050
 60 G Q S L R S G E C S L A L V G G
 TCACGGTGATGGCGTCGCCGGCGATTGCTCGAGTTCTCCGGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R

GGGCTCGGCCGGACGGCGGGCGAAGGCCTTCGGCGCGGGCGGGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTCTGGCTCGAGCGGCTCTCCG 1200
 T S F A E G A G A L V V E R L S
 5 ACGCGGAGCGCCACGGCACACCGTCTCGCCCTCGTACCGGGCTCCGCG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCGTCAACGGCTGTCCGCGCCGAACGGCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCTCATCCACCAGGCCCTCGCAACCGAAGCTACCCCCG 1350
 10 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACGCCGGTCAGGCGCACGGCACCGGCACCCGCTCGGCAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCGCAGGCCGCTCGCACGTACGGACAGGACCGGGGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 15 GCCCCTGCTCGCTCGGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCCG 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTCGCCGGATCATCAAGATGGTGCAGGCCATCCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCACACTGCCACGCCGAGGCCGCTGCCGACGTCGACTG 1600
 20 E I P P T L H A D E P S P H V D W
 GACGCCGGTGGCTCGAGCTCTGACGTGGCCGGCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCGCCGCGCTGCCGCTCGCTCGTCGGCTGAGCGGGACCG 1700
 T G R P R R A A V S S F G V S G T
 25 AACGCCACATCATCCTTGAGGCAGGCCGTCAGGACCGGCTAAAACGGGACCGGTGCA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGCAGGCCGTCAGAAGTAGGACCGGTCGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 GACCGCTCCCGCGCCGCGCCGTCAGCACCGGGCGAAGACCTCCGCTG 1850
 30 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCGTTCCCCGGAGGCCACTCGACGAGCAGATCGGGCGCCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGGCCCTATCTGACACCCGGCCGGCGTCGACCGGGCGCCGTGGCGC 1950
 R A Y L D T G P G V D R A A V A
 35 AGACACTGGCCCGCGTACGCACCTTCACCCACCGGGCGTACTGCTCGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATGGCGCTCCCCCGCGGACCAGGCCGACGAACCTCGCTT 2050
 D T V I G A P P A D Q A D E L V F
 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCATGGCGAGCAGCTAG 2100
 40 V Y S G Q G T Q H P A M G E Q L
 CCGCCGCGTTCCCCGTCTCGCGCGATCCATCAGCAGGTGTGGACCTG 2150
 A A A F P V F A R I H Q Q V W D L
 CTCGATGTGCCGATCTGGAGGTGAACGAGACCGGTTACGCCAGCCGGC 2200
 L D V P D L E V N E T G Y A Q P A
 45 CCTGTTCGCAATGCAGGTGGCTCTGTTGGCTGCTGGAAATCGTGGGTG 2250
 L F A M Q V A L F G L L E S W G
 TACGACCGGACGCCGGTATCGGCCATTGGTGGCTGAGCTTGCGGCTGCG 2300
 V R P D A V I G H S V G E L A A A
 TATGTGTCCGGGTGTGGTCGTTGGAGGATGCCACTTGGTGTGGC 2350
 50 Y V S G V W S L E D A C T L V S A
 GCGGGCTCGTCGATGCAGGCTCTGCCCGCGGGTGGGTGATGGTCGCTG 2400
 R A R L M Q A L P A G G V M V A
 TCCCGGTCTCGGAGGATGAGGCCGCCGTGCTGGGTGAGGGTGTGGAG 2450
 V P V S E D E A R A V L G E G V E
 55 ATGCCCGGGTCAACGCCCGTCGTCGGTGGTCTCTCCGGTATGAGGCC 2500
 I A A V N G P S S V V L S G D E A
 CGCCGTGCTGCAGGCCGCCGGAGGGCTGGGGAAAGTGGACGCCGCGA 2550
 A V L Q A A E G L G K W T R L A
 CCAGCCACGCCGTTCCATTCCGCCGTATGAAACCCATGCTGGAGGAGTC 2600
 60 T S H A F H S A R M E P M L E E F
 CGGGCGGTGCGCGAAGGCCCTGACCTACCGGACGCCGAGGTCTCCATGGC 2650
 R A V - A E G L T Y R T P Q V S M A
 CGTTGGTGTAGGTGACCGACCGCTGAGTACTGGGTGCGGCAGGTCCGGG 2700
 V G D Q V T T A E Y W V R Q V R

ACACGGTCCGGTTGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTTC 2750
 D T V R F G E Q V A S Y E D A V F
 GTCGAGCTGGTGCCGACCGGTCACTGGCCCCTGGTCACGGTGTGC 2800
 V E L G A D R S L A R L V D G V A
 5 GATGCTGCACGGCACCACGAAATCCAGGCCGATCGGCCCTGGCCC 2850
 M L H G D H E I Q A A I G A L A
 ACCTGTATGTCAACGGCGTACGGTCACTGGCCGCTCCCTGGCGAT 2900
 H L Y V N G V T V D W P A L L G D
 GCTCCGGCAACACGGGTGCTGGACCTTCGACATA CGCCTCCAGCACCA 2950
 10 A P A T R V L D L P T Y A F Q H Q
 GCGCTACTGGCTCGAGTCGGCTCCCCCGGCCACGGCCACTCGGCCACC 3000
 R Y W L E S A P P A T A D S G H
 CCGTCCTCGGCACCGGAGTCGCCGTCGCCGGTCGCCGGCGGGTTC 3050
 P V L G T G V A V A G S P G R V F
 15 ACGGGTCCCGTGCCC GCCGGTGC GGACCGCGCGGTGTTCATGCCGA 3100
 T G P V P A G A D R A V F I A E L
 GGC GCTCGCCGCCGACGCCACCGACTGCCACGGTCGAACAGCTCG 3150
 A L A A A D A T D C A T V E Q L
 ACGTCACCTCCGTGCCCGGGATCCGCCCGGGCAGGGCCACCGCGCAG 3200
 20 D V T S V P G G S A R G R A T A Q
 ACCTGGGTGATGAACCCGCCGACGGCGGCCGCTTCACCGTCCA 3250
 T W V D E P A A D G R R R F T V H
 CACCCCGTGC CGACGCCCGTGGACGCTGCACGCCGAGGGGGTTCTCC 3300
 T R V G D A P W T L H A E G V L
 25 GCCCCGGCCGCGTCCCCAGCCCGAACGCCGTCACCCGCTGGCCCCCG 3350
 R P G R V P Q P E A V D T A W P P
 CCGGGCGCGTCCCGCGGACGGGCTGCCCGGGCTGGCGACGCCGCGA 3400
 P G A V P A D G L P G A W R R A D
 CCAGGTCTCGTCAAGCCGAAGTCGACAGCCCTGACGGCTCGTGGCAC 3450
 30 Q V F V E A E V D S P D G F V A
 ACCCCGACCTGCTCGACGGGTCTTCTCCCGGTGGCGACGGGAGCCGC 3500
 H P D L L D A V F S A V G D G S R
 CAGCCGACCGGATGGCGCGACCTCGCGGTGCACCGCTCGACGCCACCGT 3550
 Q P T' G W R D L A V H A S D A T V
 35 GCTGCCGCCCTGCTCACCCGCCGACAGTGGTGTGGAGCTGCCG 3600
 L R A C L T R R D S G V V E L A
 CCTTCGACGGTCCGGAAATGCCGGTGTCAACCGCGAGTCGGTGACGCTG 3650
 A F D G A G M P V L T A E S V T L
 GGC GAGGTGCGCGTCCGCAAGCGGATCCGACGAGTCGGACGGTCTGCTCG 3700
 40 G E V A S A G G S D E S D G L L R
 GCTTGAGTGGTTGCCGGTGGCGGAGGCCACTACGACGGTGCCGACGAGC 3750
 L E W L P V A E A H Y D G A D E
 TGCCCGAGGGCTACACCCCTCATCACCGCACACACCCGACGACCCGAC 3800
 L P E G Y T L I T A T H P D D P D
 45 GACCCCACCAACCCCCACAACACACCCACACGACCCACACACAAACAC 3850
 D P T N P H N T P T R T H T Q T T
 ACGCGTCTCACC CGCCCTCAACACCCACCTCATCACCAACCAACACCC 3900
 R V L T A L Q H H L I T T N H T
 TCATCGTCCACACCACCGACCCCCCAGGCCGCCGTACCGGCTC 3950
 50 L I V H T T T D P P G A A V T G L
 ACCCGCACCGCACAAACGAACACCCGGCCGATCCACCTCATCGAAC 4000
 T R T A Q N E H P G R I H L I E T
 CCACCAACCCCCACACCCACTCCCCCTCACCAACTCACCAACCTCCACC 4050
 H H P H T P L P L T Q L T T L H
 55 AACCCCACCTACGCCCTACCAACACACCCCTCACACCCCCCACCTCACC 4100
 Q P H L R L T N N T L H T P H L T
 CCCATCACCAACCCACCAACACCCACACACCCACACACCCACCC 4150
 P I T T H H N T T T T P N T P P
 CCTCAACCCCAACCACGCCATCCCTCATCACCGGCCGTCCGGCACCC 4200
 60 L N P N H A I L I T G G S G T L
 CGGCATCCTCGCCGCCACCTCAACCACCCCCACACCTACCTCCTCTCC 4250
 A G I L A R H L N H P H T Y L L S
 CGCACACCACCAACCCCCACACACCCGGCACCCACATCCCCTGCGACCT 4300
 R T P P P P T T P G T H I P C D L

CACCGACCCCCACCCAAATCACCAAGCCCTCACCCACATACCACAACCCC 4350
 T D P T Q I T Q A L T H I P Q P
 TCACGGCATCTTCCACACCGCCGCCCCCTCGACGACGCCACCCCTCACC 4400
 L T G I F H T A A T L D D A T L T
 5 AACCTCACCCCCCAACACCTCACCAACCCACCCCTCAACCCAAAGCCGACGC 4450
 N L T P Q H L T T T L Q P K A D A
 CGCCTGGCACCTCCACCACCAACCCAAAACCAACCCCTCACCCACTTCG 4500
 A W H L H H H T Q N Q P L T H F
 TCCTCTACTCCAGCGCCGCCACCCCTCGGCAGCCCCGCCAAGCCAAC 4550
 10 V L Y S S A A A T L G S P G Q A N
 TACGCCGCCAACGCCCTCGACGCCCTCGCCACCCACCGCCACAC 4600
 Y A A A N A F L D A L A T H R H T
 CCAAGGACAACCCGCCACCACCATCGCCTGGGCATGTGGCACACCAACCA 4650
 Q G Q P A T T I A W G M W H T T
 15 CCACACTCACCAAGCCAACCTCACCGACAGCGACCGCAGCCATCCGCCGC 4700
 T T L T S Q L T D S D R D R I R R
 GGCGGCTTCCCTGCCGATCTGGACGACGAGGGCATGC
 G G F L P I S D D E G M

20 The *NheI-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 13 of rapamycin is shown below.

GCATGCGGCTGTACGAGGCCGACGGCGCACCGGAAGTCCC GTGGTGGTG 50
 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGCCGGACGTGCCGCTGCTGCCGGCTGCG 100
 25 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCTGCCGCCGTCCGGAAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGAGCGCGCCACGCGCTCCCTCGCGTTCG 200
 R S P C C P T T S A P T P P S R S
 30 TCCCTGGAACAGCACCGCCACCGTGTGCTGGCCACCTGGGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACTCGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCGCAACGCGCTGACCACGGCGACCGCGTACGCCCTAACGCC 350
 35 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTCGACTTCCGACGCCGCCGCGCTGCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCGCGCCCGTCGCCGGCCGACCGCGCCA 450
 D E L A G T R A P V A A R T A A
 40 CCGCGGCCGCGCACGAAACCGCTGGCGATCGTGGCATGGCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCGGGCGGGGTGCGCTGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTTCCCCGCCGGACCGCGCTGGACGTGG 600
 45 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCGACGCGATCGGCAAGACCTTCGTCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCCTCGACGGTGTGCGACCGGCTTCGACGCCGGCTTCGG 700
 H G G F L D G A T G F D A A F F G
 50 GATCAGCCCGCCGAGGCCCTGGCATGGACCCCGACGACCGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCTGGGAGGGCTTCGAAAGCGCGGGCATACCCCGACCG 800
 L E T S W E A F E S A G I T P D A
 GCGCGGGGGCAGCGACACCGGGCTGTCACTGGCGCGTCTCCCTACGGGTA 850
 55 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGCAGAGGGTCGACGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCTCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTTCG 950
 S V L S G R L S Y F Y G L E G P S
 60 GTCACGGTCGACACCGCCCTGCTCGTCGTCAGGGTCCGACACCAGGC 1000
 V T V - D T A C S S S L V A L H Q A
 AGGGCAGTCCCTGCGCTCGGCGAATGCTCGCTGCCCTGGTCGGCGGTG 1050

G Q S L R S G E C S L A L V G G
 TCACGGTATGGCGTCGCCCGGATTCTCGAGTTCTCCGGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGCGAAGGCCTCGCGCGGGCGCG 1150
 5 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTCCCTGGTGGTCGAGCGGCTCTCCG 1200
 T S F A E G A G A L V V E R L S
 ACGCGGAGCGCCACGGCACACCGTCCCTGCCCTCGTACCGGGCTCCG 1250
 D A E R H G H T V L A L V R G S A
 10 GCTAACTCCGACGGCGTCGAACGGTCTGTCGGGCCGAACGGCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAGGCCCTCGGAACCGCAAACCCCCCG 1350
 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACGCCGCTGAGGCACGGCACCCGCACCCGCTCGGCAC 1400
 15 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCGAGGCCTGCTCGCACGGTACGGACAGGACGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCTGCTGCTCGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 20 CGTCAGGGGTCCCGGGATCATCAAGATGGTGCAGGCCATCCGGCACGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCAGAGCCGTCGCCGACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 GACGCCCGGTGCCGTCGAGCTCTGACGTCGGCCCGGTGGCCGGGA 1650
 25 T A G A V E L L T S A R P W P G
 CCGGTCGCCGCCGCGCTGCCGTCTCGTCGTTGGCGTGAGCGGCACG 1700
 T G R P R R A A V S S F G V S G T
 AACGCCACATCATCCTTGAGGCAGGACCGGTCAAAACGGGACCGGTCGA 1750
 N A H I I L E A G P V K T G P V E
 30 GGCAGGAGCGATCGAGGCAGGACCGGTCGAAGTAGGACCGGTCGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 GACCGCTCCCGCGCCGCGCTCACCGCCGCGTCAGCACCGGGCGAAGACCTCCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCGTCTCCCGGAGGCACTCGACGAGCAGATCGGGCGCCT 1900
 35 L V S A R S P E A L D E Q I G R L
 GCGGCCCTATCTGACACCCGGCCGGCGTCGACCGGGCGCGTGGCGC 1950
 R A Y L D T G P G V D R A A V A
 AGACACTGGCCCGCGTACGCACCTCACCCACCGGGCGTACTGCTCGG 2000
 Q T L A R R T H F T H R A V L L G
 40 GACACCGTCATGGCGCTCCCCCGCGGACCAGGCCGACGAACTCGCTT 2050
 D T V I G A P P A D Q A D E L V F
 CGTCGACTCCGGTCAGGGCACCCAGCATCCCGCATGGCGAGCAGCTAG 2100
 V Y S G Q G T Q H P A M G E Q L
 CCGATTGTCGGTGGTTCGCGAGCGGATGCCGAGTGTGCGGCCGGCG 2150
 45 A D S S V V F A E R M A E C A A A
 TTGCGCGAGTTCGTGGACTGGATCTGTTCACGGTTCTGGATGATCCGGC 2200
 L R E F V D W D L F T V L D D P A
 GGTGGTGGACCGGGTTGATGTGGTCAGCCCGCTCTGGCGATGATGG 2250
 V V D R V D V V Q P A S W A M M
 50 TTTCCCTGGCCCGGGTGTGGCAGGGCCGGTGTGCGGCCGGATGGCTG 2300
 V S L A A V W Q A A G V R P D A V
 ATCGGCCATTCCGCAAGGGTGAAGATGCCGCAGCTGTGCGGCCGGT 2350
 I G H S Q G E I A A A C V A G A V
 GTCACACTACGCGATGCCGCCGGATCGTGCACCTTGCACGCCAGGCATCG 2400
 55 S L R D A A R I V T L R S Q A I
 CCCGGGGCTGGCGGGCCGGGGCGATGGCATCCGTCGCCCTGCCCG 2450
 A R G L A G R G A M A S V A L P A
 CAGGATGTCGAGCTGGTCGACGGGGCTGGATGCCGCCACAACGGGCC 2500
 Q D V E L V D G A W I A A H N G P
 60 CGCCTCCACCGTGATGCCGGCACCCCGGAAGCGGTGAGCATGTCCTCA 2550
 A S T V I A G T P E A V D H V L
 CCGCTCATGAGGCACAAGGGGTGCGGGTGCAGGATCACCGTCGACTAT 2600
 T A H E A Q G V R V R R I T V D Y
 GCCTCGCACACCCCGCACGTCGAGCTGATCCCGACGAACACTCGACAT 2650

A S H T P H V E L I R D E L L D I
 CACTAGCGACAGCAGCTCGCAGACCCCGCTGCGCTGGCTGTCGACCG 2700
 T S D S S S Q T P L V P W L S T
 TGGACGGCACCTGGGTCGACAGCCCGCTGGACGGGAGTACTGGTACCGG 2750
 5 V D G T W V D S P L D G E Y W Y R
 AACCTGCGTAACCGGTGGTTCCACCCCGCCGTACGCCAGTTGCAGGC 2800
 N L R E P V G F H P A V S Q L Q A
 CCAGGGCGACACCGTGTTCGTCGAGGTCAAGCAGCCAGCCGGTGTG 2850
 Q G D T V F V E V S A S P V L L
 10 AGGGATGGACGACGATGTCGTACCGTTGCCACGCTGCGTCGTGACGAC 2900
 Q A M D D D V V T V A T L R R D D
 GGCAGGCCACCCGGATGCTCACCGCCCTGGCACAGGCCATGTCCACGG 2950
 G D A T R M L T A L A Q A Y V H G
 CGTACCGTCGACTGGCCCGCCATCTCGGCACCACCAACCCGGGTAC 3000
 15 V T V D W P A I L G T T T R V
 TGGACCTTCCGACCTACGCCCTCAAACACAGCGGTACTGGCTCGAGTCG 3050
 L D L P T Y A F Q H Q R Y W L E S
 GCTCCCCGGCAGGGCCACTCGGCCACCCCGTCTCGGCACCGGAGT 3100
 A P P A T A D S G H P V L G T G V
 20 CGCGTGCCTGGTGCCTGGGGCGGGTGTTCACGGTCCCGTGCCCG 3150
 A V A G S P G R V F T G P V P A
 GTGCGGACCGCGCGGTGTTCATCGCCGAACACTGGCGCTGCCGCCGAC 3200
 G A D R A V F I A E L A L A A A D
 GCCACCGACTGCGCCACGGTCGAACAGCTCGACGTACCTCCGTGCCGG 3250
 25 A T D C A T V E Q L D V T S V P G
 CGGATCCGCCCGCGGCAGGGCCACCGCGCAGACCTGGTCGATGAACCCG 3300
 G S A R G R A T A Q T W V D E P
 CCGCCGACGGCGGCCGCTTACCGTCCACACCGCGTCGGCGACGCC 3350
 30 A A D G R R R F T V H T R V G D A
 CCGTGGACGCTGCACGCCGAGGGGTTCTCCGCCCGCGCGTGCCCCA 3400
 P W T L H A E G V L R P G R V P Q
 GCCCGAAGCCGTCGACACCGCCTGGCCCCCGCCGGCGCGGTGCCCCGG 3450
 P E A V D T A W P P P G A V P A
 ACGGGCTGCCGGGGCGTGGCGACGGCGGACAGGTCTCGTCGAAGCC 3500
 35 D G L P G A W R R A D Q V F V E A
 GAAGTCGACAGCCCTGACGGCTTCGTCGGCACACCCGACCTGCTCGACGC 3550
 E V D S P D G F V A H P D L L D A
 GGTCTTCTCCGCCGGTGGCGACGGAGGCCAGCCGACCGGATGGCGCG 3600
 V F S A V G D G S R Q P T G W R
 40 ACCTCGCGGTGCACCGCTGGACGCCACCGTGTGCGCGCCTGCCCTCACC 3650
 D L A V H A S D A T V L R A C L T
 CGCCGCGACAGTGGTGTGGAGCTGCCGCTTCGACGGTGCCGGAAT 3700
 R R D S G V V E L A A F D G A G M
 GCCGGTGTCAACCGCGGAGTCGGTACGCTGGCGAGGTGCGCGCAG 3750
 45 P V L T A E S V T L G E V A S A
 GCGGATCCGACGAGTCGGACGGTCTGCTTGGCTTGAGTGGTGCCGGT 3800
 G G S D E S D G L L R E W L P V
 GCGGAGGCCACTACGACGGTGGCGACGAGCTGCCGAGGGCTACACCCCT 3850
 50 A E A H Y D G A D E L P E G Y T L
 CATCACCGCCACACACCCGACGACCCGACGACCCCACCAACCCACCA 3900
 I T A T H P D D P D D P T N P H
 ACACACCCACACGCACCCACACACAAACACACCGCTCCACCGCCCTC 3950
 N T P T R T H T Q T T R V L T A L
 CAACACCACTCATCACCAACACACCCCTCATCGTCCACACCACAC 4000
 55 Q H H L I T T N H T L I V H T T T
 CGACCCCCCAG3CGCCGCCGTACCGGCCCTCACCCGCACCGCACAAAAGC 4050
 D P P G A A V T G L T R T A Q N
 AACACCCCCGGCCGACCCACCTCATCGAAACCCACCCACACCCCA 4100
 E H P G R I H L I E T H H P H T P
 60 CTCCCCCTCACCCAACTCACCAACCCCTCCACCAACCCACCTACGCCCTCAC 4150
 L P L T Q L T T L H Q P H L R L T
 CAACAAACACCCCTCACACCCCCCACCTCACCCCATCACCAACCCACCA 4200
 N N T L H T P H L T P I T T H H
 ACACCAACCAACCAACCCCCAACACCCACCCCTCAACCCCAACCAACGCC 4250

```

N T T T T P N T P P L N P N H A
ATCCTCATCACCGCGGCTCCGGCACCCCTGCCGGCATCCCGCC 4300
I L I T G G S G T L A G I L A R H
CCTCAACCACCCCCACACCTACCTCTCCCGCACACCACCCCCA 4350
5 L N H P H T Y L L S R T P P P P
CCACACCCGGCACCCACATCCCCTGCGACCTCACCGACCCCACCCAAATC 4400
T T P G T H I P C D L T D P T Q I
ACCCAAGCCCTCACCCACATACCAACCCCTACCGGCATCTCCACAC 4450
T Q A L T H I P Q P L T G I F H T
10 CGCCGCCACCCCTGACGACGCCACCCCTACCAACCTCACCCCCCAACACC 4500
A A T L D D A T L T N L T P Q H
TCACCACCACCCCTCAACCAAAGCCGACGCCCTGGCACCTCCACCAC 4550
L T T T L Q P K A D A A W H L H H
CACACCCAAAACCAACCCCTCACCCACTTCGTCTACTCCAGGCCGC 4600
15 H T Q N Q P L T H F V L Y S S A A
CGCCACCCCTCGCAGCCCCGGCAAGCCAACACTACGCCGCCAACGCCT 4650
A T L G S P G Q A N Y A A A N A
TCCTCGACGCCCTGCCACCCACCGCCACACCAAGGACAACCCGCCACC 4700
20 F L D A L A T H R H T Q G Q P A T
ACCATCGCCTGGGCATGTGGCACACCACCCACTCACCAACT 4750
T I A W G M W H T T T T L T S Q L
CACCGACAGCGACCGCGACCGCATCCGCCGCCGGCTTCCCTGCCGATCT 4800
T D S D R D R I R R G G F L P I
25 CGGACGACGAGGGCATGC
S D D E G M

```

Example 3Recombinant PKS Genes for 13-desmethoxy FK-506 and FK-520

The present invention provides a variety of recombinant PKS genes in addition to those described in Examples 1 and 2 for producing 13-desmethoxy FK-506 and FK-520 compounds. This Example provides the construction protocols for recombinant FK-520 and FK-506 (from *Streptomyces* sp. MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference) PKS genes in which the module 8 AT coding sequences have been replaced by either the *rapAT3* (the AT domain from module 3 of the rapamycin PKS), *rapAT12*, *eryAT1* (the AT domain from module 1 of the erythromycin (DEBS) PKS), or *eryAT2* coding sequences. Each of these constructs provides a PKS that produces the 13-desmethoxy-13-methyl derivative, except for the *rapAT12* replacement, which provides the 13-desmethoxy derivative, i.e., it has a hydrogen where the other derivatives have methyl.

Figure 7 shows the process used to generate the AT replacement constructs. First, a fragment of ~4.5 kb containing module 8 coding sequences from the FK-520 cluster of ATCC 14891 was cloned using the convenient restriction sites *SacI* and *SphI* (Step A in Figure 7). The choice of restriction sites used to clone a 4.0 - 4.5 kb fragment comprising module 8 coding sequences from other FK-520 or FK-506 clusters can be different depending on the DNA sequence, but the overall scheme is identical. The unique *SacI* and *SphI* restriction sites at the ends of the FK-520 module 8 fragment were then changed to unique *Bgl* II and *Nsi*I sites by ligation to synthetic linkers (described in

the preceding Examples, see Step B of Figure 7). Fragments containing sequences 5' and 3' of the AT8 sequences were then amplified using primers, described above, that introduced either an *AvrII* site or an *NheI* site at two different KS/AT boundaries and an *XhoI* site at the AT/DH boundary (Step C of Figure 7). Heterologous AT domains from 5 the rapamycin and erythromycin gene clusters were amplified using primers, as described above, that introduced the same sites as just described (Step D of Figure 7). The fragments were ligated to give hybrid modules with in-frame fusions at the KS/AT and AT/DH boundaries (Step E of Figure 7). Finally, these hybrid modules were ligated into the *BamHI* and *PstI* sites of the KCS15 vector. The resulting recombinant phage 10 were used to transform the FK-506 and FK-520 producer strains to yield the desired recombinant cells, as described in the preceding Examples.

The following table shows the location and sequences surrounding the engineered site of each of the heterologous AT domains employed. The FK-506 hybrid construct was used as a control for the FK-520 recombinant cells produced, and a similar FK-520 15 hybrid construct was used as a control for the FK-506 recombinant cells.

Heterologous AT	Enzyme	Location of Engineered Site
FK-506 AT8 (hydroxymalonyl)	<i>AvrII</i>	GGCCGT <u>ccgcgc</u> CGTGC GGGGT CTCGTCGTTC G R P R R A A V S S F
	<i>NheI</i>	ACCCAGCATCCC CG GATGGGTGAGCG <u>gctcgcc</u> C T Q H P A M G E R L A
	<i>XhoI</i>	TACGCC TT CCAGCGGCC CT ACTGG <u>atcgag</u> Y A F Q R R P Y W I E
rapamycin AT3 (methylmalonyl)	<i>AvrII</i>	GACCGG <u>ccccgt</u> CGGGC GGGGT GT CGT CC CTTC D R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGGCTGGGATGGGCAGTGC <u>cctgcqG</u> W Q W L G M G S A L R
	<i>XhoI</i>	TACGCC TT CCAACACCAGCGGTACTGG <u>gtcgag</u> Y A F Q H Q R Y W V E
rapamycin AT12 (malonyl)	<i>AvrII</i>	GGCCGA <u>gcgcgc</u> CGGGCAGGCGTGT CGT CC CTTC G R A R R A G V S S F
	<i>NheI</i>	TCGCAGCGT GCT GGCATGGGTGAGGA <u>actggcC</u> S Q R A G M G E E L A
	<i>XhoI</i>	TACGCC TT CCAGCACCA GCGT ACTGG <u>ctcgag</u> Y A F Q H Q R Y W L E
DEBS AT1 (methylmalonyl)	<i>AvrII</i>	GCGCGA <u>ccgcgc</u> CGGGC GGGGT CTCGTCGTTC A R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGGGCGGGCATGGCC GT CGA <u>acctgctC</u> W Q W A G M A V D L L
	<i>XhoI</i>	TACCCGTTCCAGCGC GAG CGCGT CT GG <u>ctcgaa</u> Y P F Q R E R V W L E
DEBS AT2 (methylmalonyl)	<i>AvrII</i>	GACGGG <u>gtgcgc</u> CGGGCAGGTGT CG GTTC D G V R R A G V S A F
		GCCCAGTGGGAAGGCATGGCGCGGGAgttgttG

	<i>NheI</i>	A Q W E G M A R E L L TATCCCTTCAGGGCAAGCGGTTCTGG <u>ctgctg</u> Y P F Q G K R F W L L
	<i>XbaI</i>	

The sequences shown below provide the location of the KS/AT boundaries chosen in the FK-520 module 8 coding sequences. Regions where *AvrII* and *NheI* sites were engineered are indicated by lower case and underlining.

5 CCGCGCCGTCGAACTGCTGACGTCGGCCCGGCGTGGCCCCGAGACCGACCGccacqqC
A G A V E L L T S A R P W P E T D R P R
GTGCCGCGTCCTCGTTCGGGTGAGCGGCACCAACGCCACGTCATCCTGGAGGCCG
R A A V S S F G V S G T N A H V I L E A
GACCGGTAACGGAGACGCCCGCGCCATCGCCTCCGGTGACCTTCCCTGCTGGTGTCGG
G P V T E T P A A S P S G D L P L L V S
10 CACGCTCACCGGAAGCGCTCGACGAGCAGATCCGCCGACTGCGCCTACCTGGACACCA
A R S P E A L D E Q I R R L R A Y L D T
CCCCGGACGTCGACCGGGTGGCCGTGGCACAGACGCTGGCCGGCGCACACACTCGCCC
T P D V D R V A V A Q T L A R R T H F A
ACCGCGGCGTGTGCTCGGTGACACCGTCATACCACACACCCCCCGCGGACCGGGCCGACG
15 H R A V L L G D T V I T T P P A D R P D
AACTCGTCTTCGTCTACTCCGGCCAGGGCACCAGCATCCCGCGATGGGCGAGCAgctcq
E L V F V Y S G Q G T Q H P A M G E Q L
CCGGCGCCCATCCCGTGTTCGCGACGCCTGGCATGAAGCGCTCCGCCCCTGACAACC
A A A H P V F A D A W H E A L R R L D N
20

The sequences shown below provide the location of the AT/DH boundary chosen in the FK-520 module 8 coding sequences. The region where an *XhoI* site was engineered is indicated by lower case and underlining.

25 TCCTCGGGGCTGGGTCACGGCACGACGCGGATGTGCCCCGGTACCGCTTCCAACGGCGGC
I L G A G S R H D A D V P A Y A F Q R R
ACTACTGGAtcqaqTCGGCACGCCGGCGCATCCGACGCGGGCCACCCGGCTGGGCT
H Y W I E S A R P A A S D A G H P V L G

The sequences shown below provide the location of the KS/AT boundaries chosen in the FK-506 module 8 coding sequences. Regions where *AvrII* and *NheI* sites were engineered are indicated by lower case and underlining.

30 TCGGCCAGGCCGTGGCCCGGACCGGCCGTccgcgcCGTCGCCGGTCTCGTCGTTCGGG
S A R P W P R T G R P R R A A V S S F G
GTGAGCGGCACCAACGCCACATCATCCTGGAGGCCGACCCAGCAGGAGGCCGTCG
35 V S G T N A H I I L E A G P D Q E E P S
GCAGAACCCGGCCGGTGACCTCCCGCTGCTGTCGGACGGTCCCCGGAGGCACTGGAC
A E P A G D L P L L V S A R S P E A L D
GAGCAGATCGGGCGCTGCGCGACTATCTCGACGCCCCCCCGCGTGGACCTGGCGGCC
E Q I G R L R D Y L D A A P G V D L A A
40 GTGGCGGGGACACTGGCCACCGGTACGCACTTCTCCCACCGCGCCGTACTGCTCGGTGAC
V A R T L A T R T H F S H R A V L L G D
ACCGTCATCACCGCTCCCCCGTGGAACAGCCGGCGAGCTCGTCTCGTACTCGGGA
T V I T A P P V E Q P G E L V F V Y S G
CAGGGCACCCAGCATCCCGCGATGGGTGAGCGcgctcqCGCAGCCCTCCCGGTGTTCGCC
45 Q G T Q H P A M G E R L A A A F P V F A
GACCCGGACGTACCCGCCTACGCCTCCAGCCGGCCCTACTGGATCGAGTCCCGCCCG
D P D V P A Y A F Q R R P Y W I E S A P

The sequences shown below provide the location of the AT/DH boundary chosen in the FK-506 module 8 coding sequences. The region where an *XhoI* site was engineered is indicated by lower case and underlining.

GACCCGGACGTACCCGCCTACGCCTCCAGCCGGCCCTACTGGAtcqaqTCCCGCGCCG
D P D V P A Y A F Q R R P Y W I E S A P

Example 4Replacement of Methoxyl with Hydrogen or Methyl at C-15 of FK-506 and FK-520

The methods and reagents of the present invention also provide novel FK-506 and FK-520 derivatives in which the methoxy group at C-15 is replaced by a hydrogen or methyl. These derivatives are produced in recombinant host cells of the invention that express recombinant PKS enzymes that produce the derivatives. These recombinant PKS enzymes are prepared in accordance with the methodology of Examples 1 and 2, with the exception that AT domain of module 7, instead of module 8, is replaced. Moreover, the present invention provides recombinant PKS enzymes in which the AT domains of both modules 7 and 8 have been changed. The table below summarizes the various compounds provided by the present invention.

	Compound	C-13	C-15	Derivative Provided
15	FK-506	hydrogen	hydrogen	13, 15-didesmethoxy-FK-506
	FK-506	hydrogen	methoxy	13-desmethoxy-FK-506
	FK-506	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-506
	FK-506	methoxy	hydrogen	15-desmethoxy-FK-506
	FK-506	methoxy	methoxy	Original Compound -- FK-506
20	FK-506	methoxy	methyl	15-desmethoxy-15-methyl-FK-506
	FK-506	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-506
	FK-506	methyl	methoxy	13-desmethoxy-13-methyl-FK-506
	FK-506	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-506
	FK-520	hydrogen	hydrogen	13, 15-didesmethoxy FK-520
25	FK-520	hydrogen	methoxy	13-desmethoxy FK-520
	FK-520	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-520
	FK-520	methoxy	hydrogen	15-desmethoxy-FK-520
	FK-520	methoxy	methoxy	Original Compound -- FK-520
	FK-520	methoxy	methyl	15-desmethoxy-15-methyl-FK-520
30	FK-520	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-520
	FK-520	methyl	methoxy	13-desmethoxy-13-methyl-FK-520
	FK-520	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-520

Example 5Replacement of Methoxyl with Ethyl at C-13 and/or C-15 of FK-506 and FK-520

The present invention also provides novel FK-506 and FK-520 derivative compounds in which the methoxy groups at either or both the C-13 and C-15 positions are instead ethyl groups. These compounds are produced by novel PKS enzymes of the invention in which the AT domains of modules 8 and/or 7 are converted to ethylmalonyl specific AT domains by modification of the PKS gene that encodes the module.

Ethylmalonyl specific AT domain coding sequences can be obtained from, for example, the FK-520 PKS genes, the niddamycin PKS genes, and the tylosin PKS genes. The novel PKS genes of the invention include not only those in which either or both of the AT domains of modules 7 and 8 have been converted to ethylmalonyl specific AT domains but also those in which one of the modules is converted to an ethylmalonyl specific AT domain and the other is converted to a malonyl specific or a methylmalonyl specific AT domain.

Example 6

Neurotrophic Compounds

The compounds described in Examples 1 - 4, inclusive have immunosuppressant activity and can be employed as immunosuppressants in a manner and in formulations similar to those employed for FK-506. The compounds of the invention are generally effective for the prevention of organ rejection in patients receiving organ transplants and in particular can be used for immunosuppression following orthotopic liver transplantation. These compounds also have pharmacokinetic properties and metabolism that are more advantageous for certain applications relative to those of FK-506 or FK-520. These compounds are also neurotrophic; however, for use as neurotrophins, it is desirable to modify the compounds to diminish or abolish their immunosuppressant activity. This can be readily accomplished by hydroxylating the compounds at the C-18 position using established chemical methodology or novel FK-520 PKS genes provided by the present invention.

Thus, in one aspect, the present invention provides a method for stimulating nerve growth that comprises administering a therapeutically effective dose of 18-hydroxy-FK-520. In another embodiment, the compound administered is a C-18,20-dihydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18-hydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18,20-dihydroxy-FK-520 derivative. In other embodiments, the compounds are the corresponding analogs of FK-506. The 18-hydroxy compounds of the invention

can be prepared chemically, as described in U.S. Patent No. 5,189,042, incorporated herein by reference, or by fermentation of a recombinant host cell provided by the present invention that expresses a recombinant PKS in which the module 5 DH domain has been deleted or rendered non-functional.

5 The chemical methodology is as follows. A compound of the invention (~200 mg) is dissolved in 3 mL of dry methylene chloride and added to 45 μ L of 2,6-lutidine, and the mixture stirred at room temperature. After 10 minutes, tert-butyldimethylsilyl trifluoromethanesulfonate (64 μ L) is added by syringe. After 15 minutes, the reaction mixture is diluted with ethyl acetate, washed with saturated bicarbonate, washed with
10 brine, and the organic phase dried over magnesium sulfate. Removal of solvent *in vacuo* and flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% methanol) gives the protected compound, which is dissolved in 95% ethanol (2.2 mL) and to which is added 53 μ L of pyridine, followed by selenium dioxide (58 mg). The flask is fitted with a water condenser and heated to 70°C on a mantle. After 20 hours, the mixture is
15 cooled to room temperature, filtered through diatomaceous earth, and the filtrate poured into a saturated sodium bicarbonate solution. This is extracted with ethyl acetate, and the organic phase is washed with brine and dried over magnesium sulfate. The solution is concentrated and purified by flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% methanol) to give the protected 18-hydroxy compound. This compound is
20 dissolved in acetonitrile and treated with aqueous HF to remove the protecting groups. After dilution with ethyl acetate, the mixture is washed with saturated bicarbonate and brine, dried over magnesium sulfate, filtered, and evaporated to yield the 18-hydroxy compound. Thus, the present invention provides the C-18-hydroxyl derivatives of the compounds described in Examples 1 - 4.

25 Those of skill in the art will recognize that other suitable chemical procedures can be used to prepare the novel 18-hydroxy compounds of the invention. See, e.g., Kawai *et al.*, Jan. 1993, Structure-activity profiles of macrolactam immunosuppressant FK-506 analogues, *FEBS Letters* 316(2): 107-113, incorporated herein by reference. These methods can be used to prepare both the C18-[*S*]-OH and C18-[*R*]-OH enantiomers, with
30 the *R* enantiomer showing a somewhat lower IC₅₀, which may be preferred in some applications. See Kawai *et al.*, *supra*. Another preferred protocol is described in Umbreit and Sharpless, 1977, *JACS* 99(16): 1526-28, although it may be preferable to use 30 equivalents each of SeO₂ and t-BuOOH rather than the 0.02 and 3-4 equivalents, respectively, described in that reference.

All scientific and patent publications referenced herein are hereby incorporated by reference. The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments, that the foregoing description and example is for purposes of illustration and not limitation of the following claims.

Claims

1. An isolated nucleic acid that encodes a CoA ligase, a non-ribosomal peptide synthetase, or a domain of an extender module of a polyketide synthase enzyme that synthesizes FK-520.
5
2. The isolated nucleic acid of claim 1 that encodes an extender module, said module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
- 10 3. The isolated nucleic acid of claim 1 that encodes an open reading frame, said open reading frame comprising coding sequences for two or more extender modules, each extender module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
- 15 4. The isolated nucleic acid of claim 1 that encodes a gene cluster, said gene cluster comprising two or more open reading frames, each of said open reading frames comprising coding sequences for two or more extender modules, each of said extender modules comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
20
5. The isolated nucleic acid of claim 2, wherein at least one of said domains is a domain of a module of a non-FK-520 polyketide synthase.
- 25 6. The isolated nucleic acid of claim 1, wherein said nucleic acid is a recombinant vector capable of replication in or integration into the chromosome of a host cell.
- 30 7. The isolated nucleic acid of claim 6 that is selected from the group consisting of cosmid pKOS034-120, cosmid pKOS034-124, cosmid pKOS065-M27, and cosmid pKOS065-M21.
8. The isolated nucleic acid of claim 5, wherein said non-FK-520 polyketide synthase is rapamycin polyketide synthase, FK-506 polyketide synthase, or erythromycin polyketide synthase.
35

9. A method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector of claim 6, and culturing said host cell under conditions such that said polyketide synthase is produced and catalyzes synthesis of said polyketide.

5

10. The method of claim 9, wherein said host cell is a *Streptomyces* host cell.

11. The method of claim 9, wherein said polyketide is selected from the group consisting of FK-520, 13-desmethoxy-FK-520, and 13-desmethoxy-FK-506.

10

12. A recombinant host cell that expresses a recombinant polyketide synthase selected from the group consisting of: (i) an FK-520 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-520 polyketide synthase; (ii) an FK-506 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-506 polyketide synthase; (iii) an FK-520 polyketide synthase in which at least one DH domain has been deleted; (iv) an FK-506 polyketide synthase in which at least one DH domain has been deleted.

15

13. The recombinant host cell of claim 12 that expresses an FK-520 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

20

14. The recombinant host cell of claim 12 that expresses an FK-506 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

25

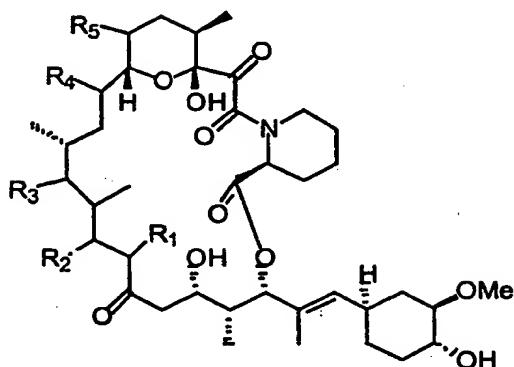
15. The recombinant host cell of claim 13, wherein a DH domain of module 5 or module 6 has been deleted.

30

16. The recombinant host cell of claim 14, wherein a DH domain of module 5 or module 6 has been deleted.

17. A recombinant host cell that comprises recombinant genes coding for enzymes sufficient for synthesis of ethylmalonyl CoA or 2-hydroxymalonyl CoA.

35

18. A polyketide having the structure

5 wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506.

10

19. The polyketide of claim 18 that is 13-desmethoxy-FK-506.**20. The polyketide of claim 18 that is 13-desmethoxy-18-hydroxy-FK-520.**

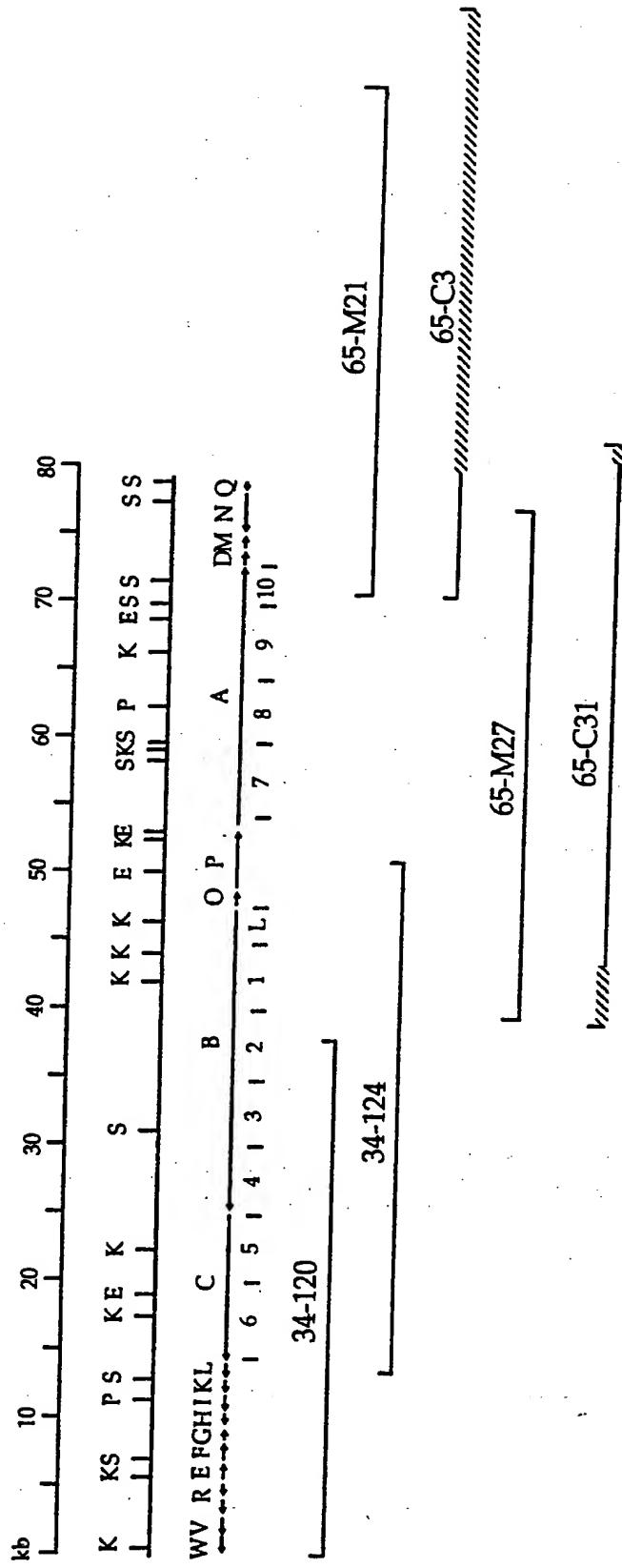


Figure 1

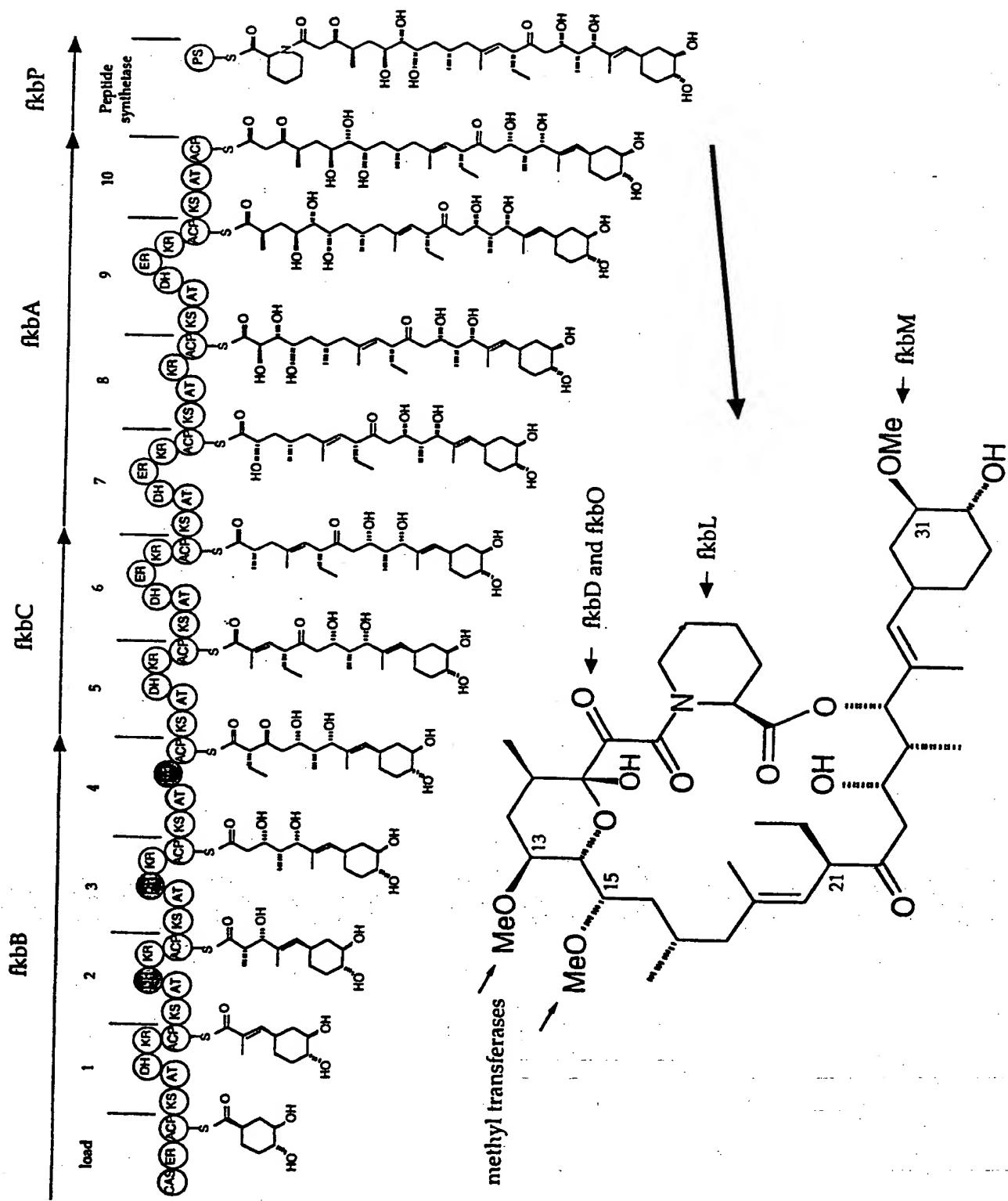


Figure 2

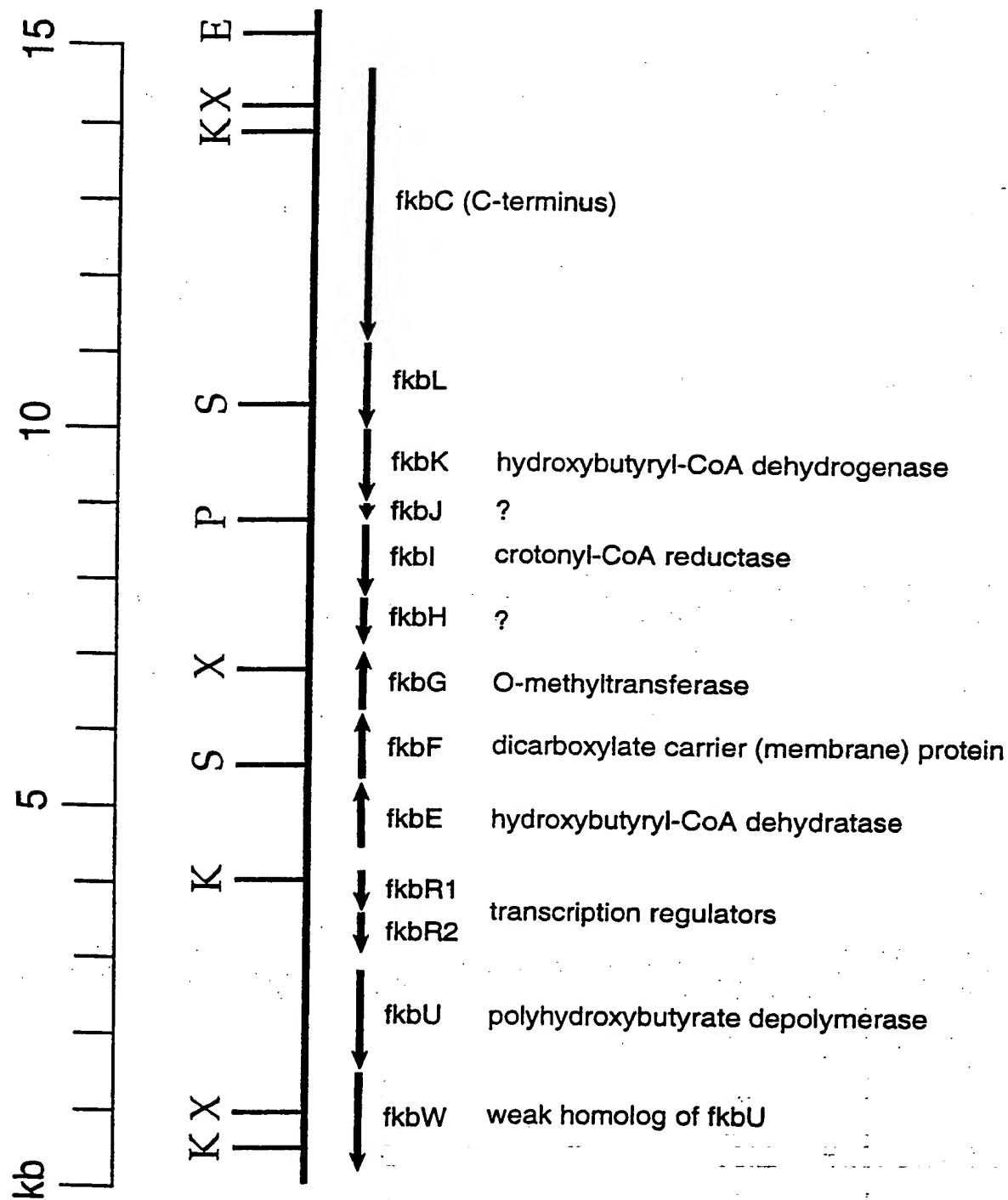


Figure 3

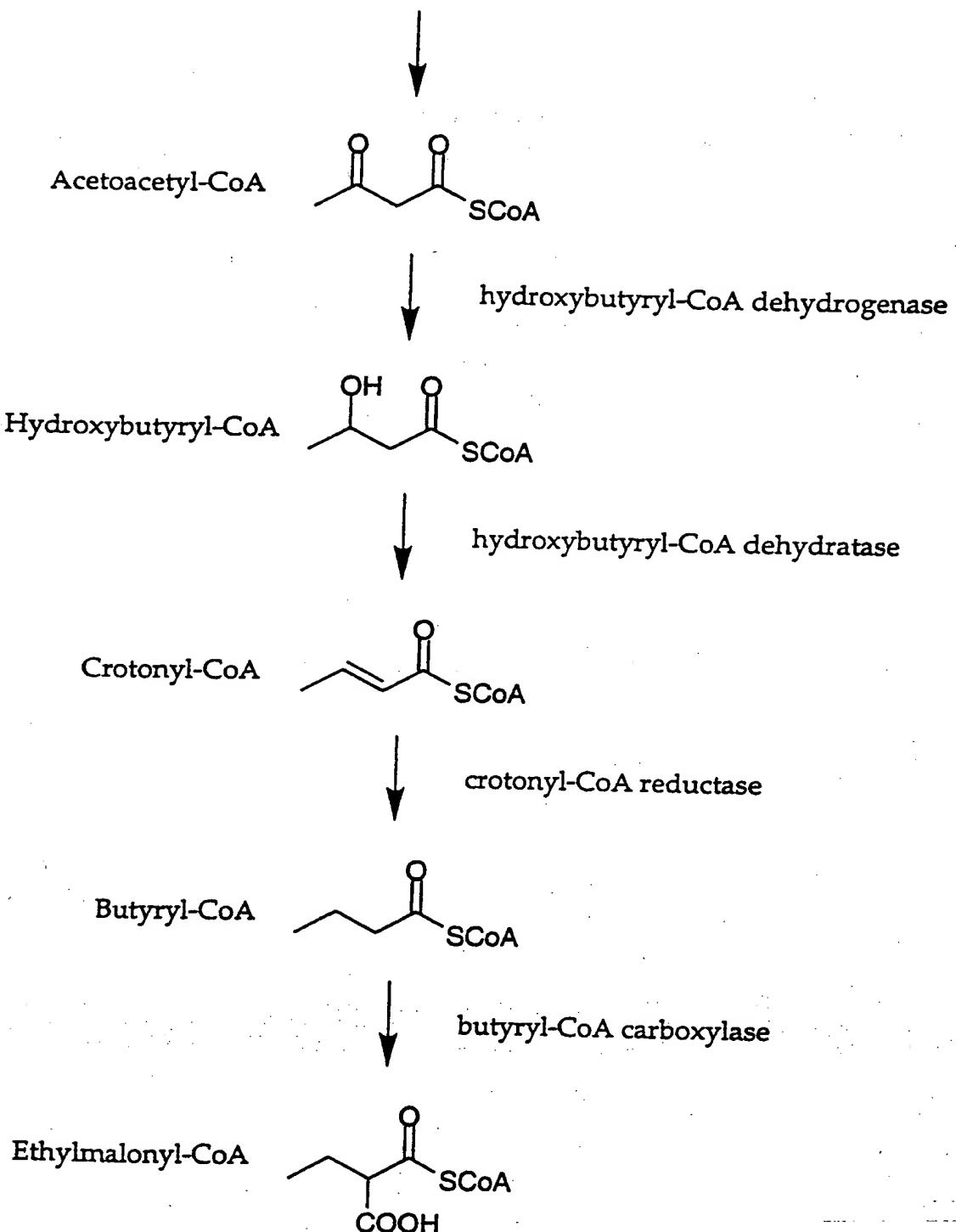


Figure 4

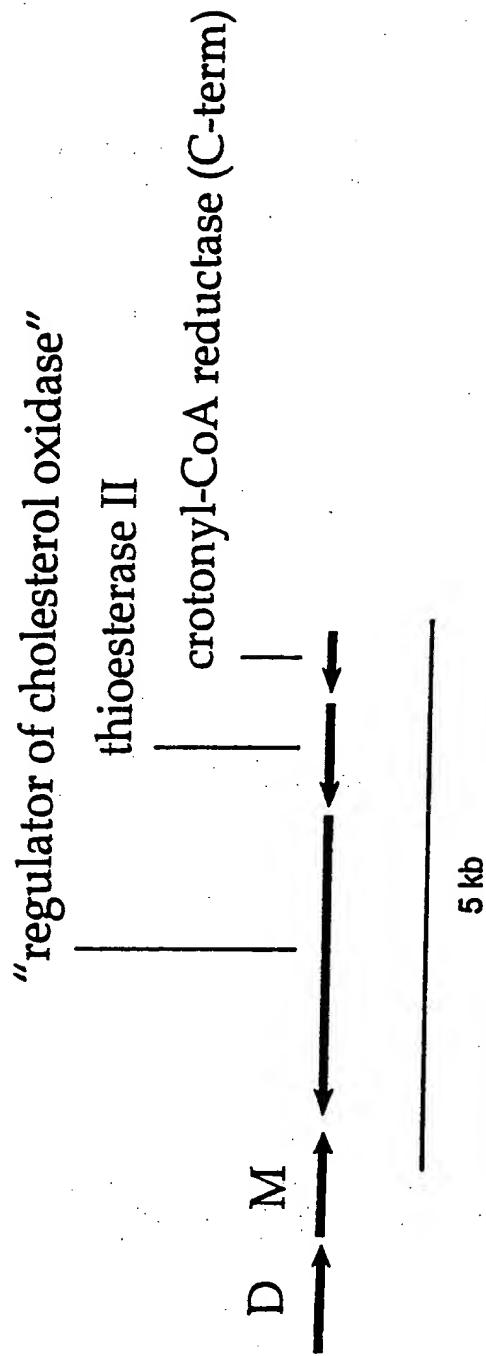


Figure 5

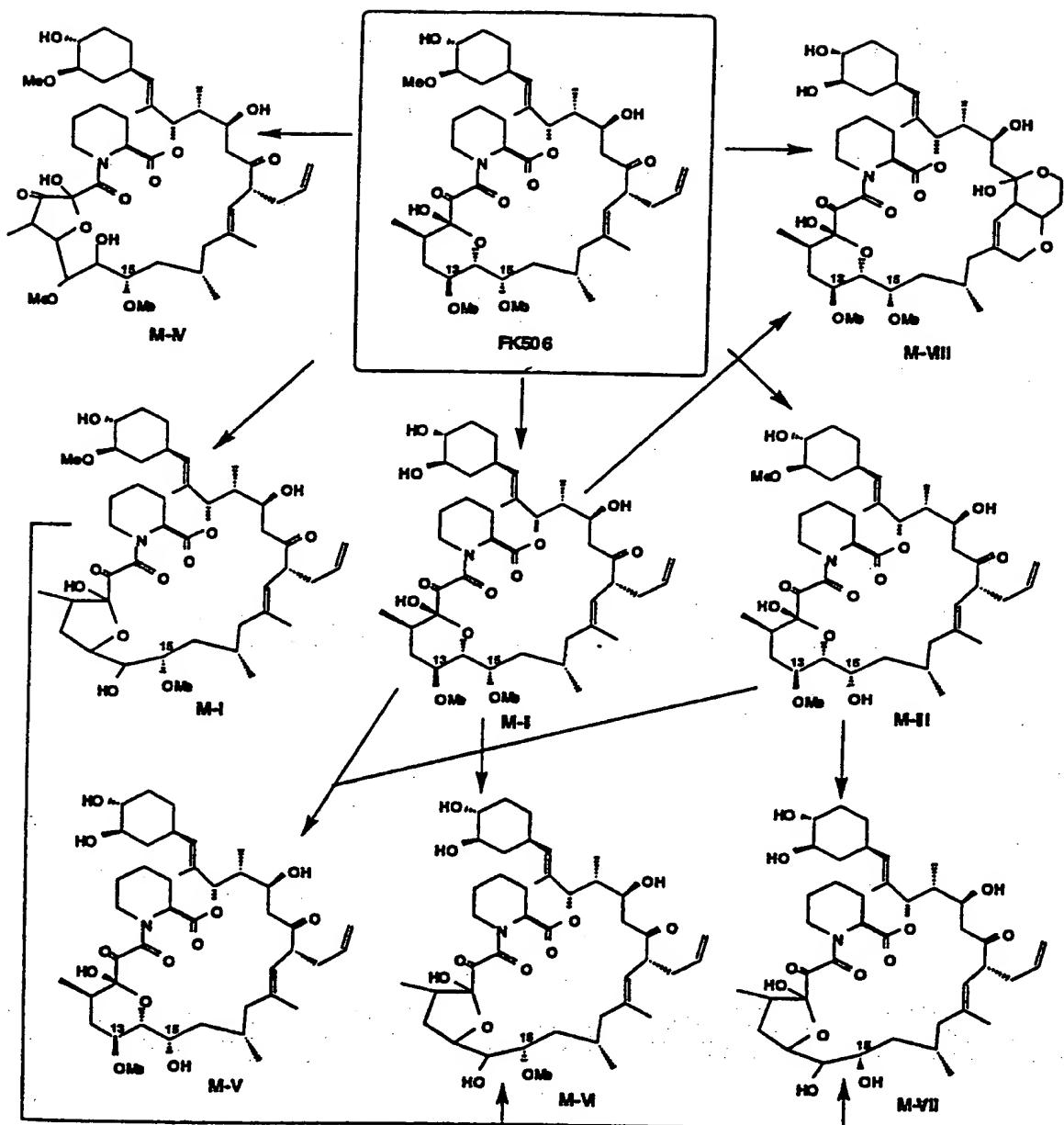


Figure 6

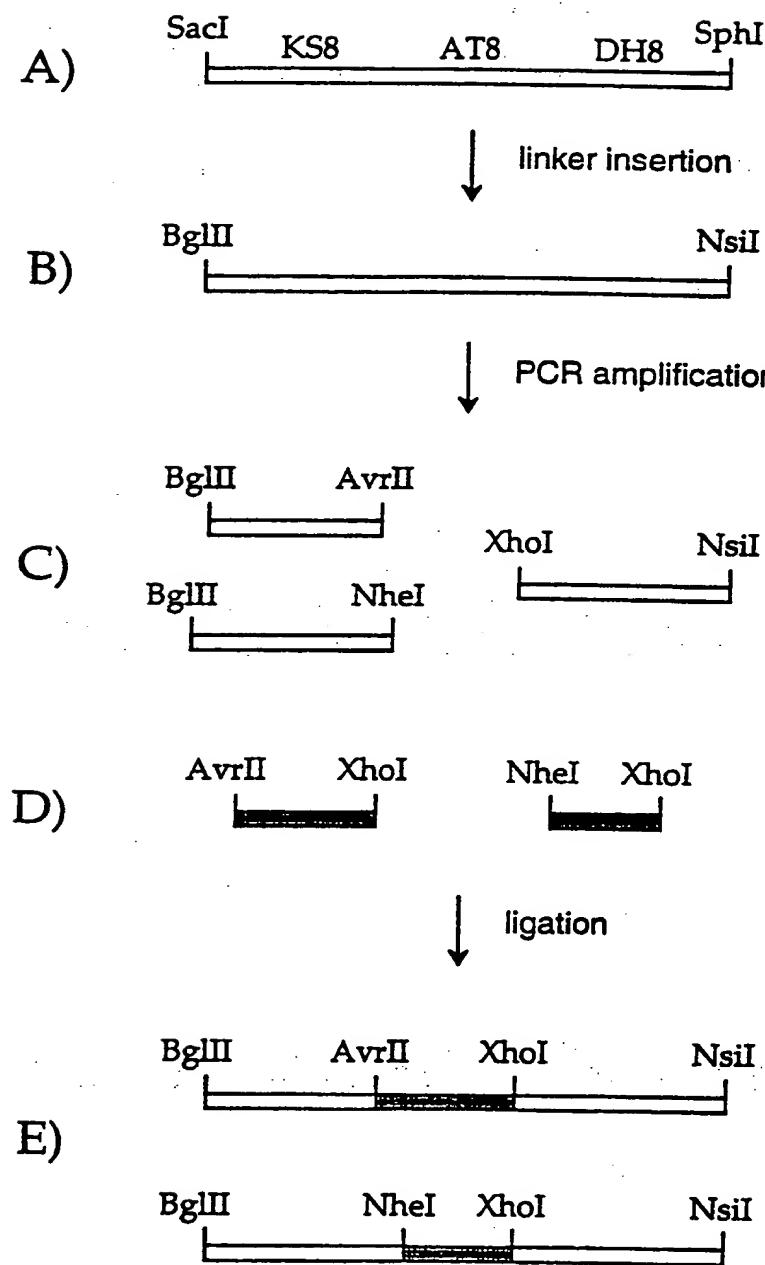


Figure 7

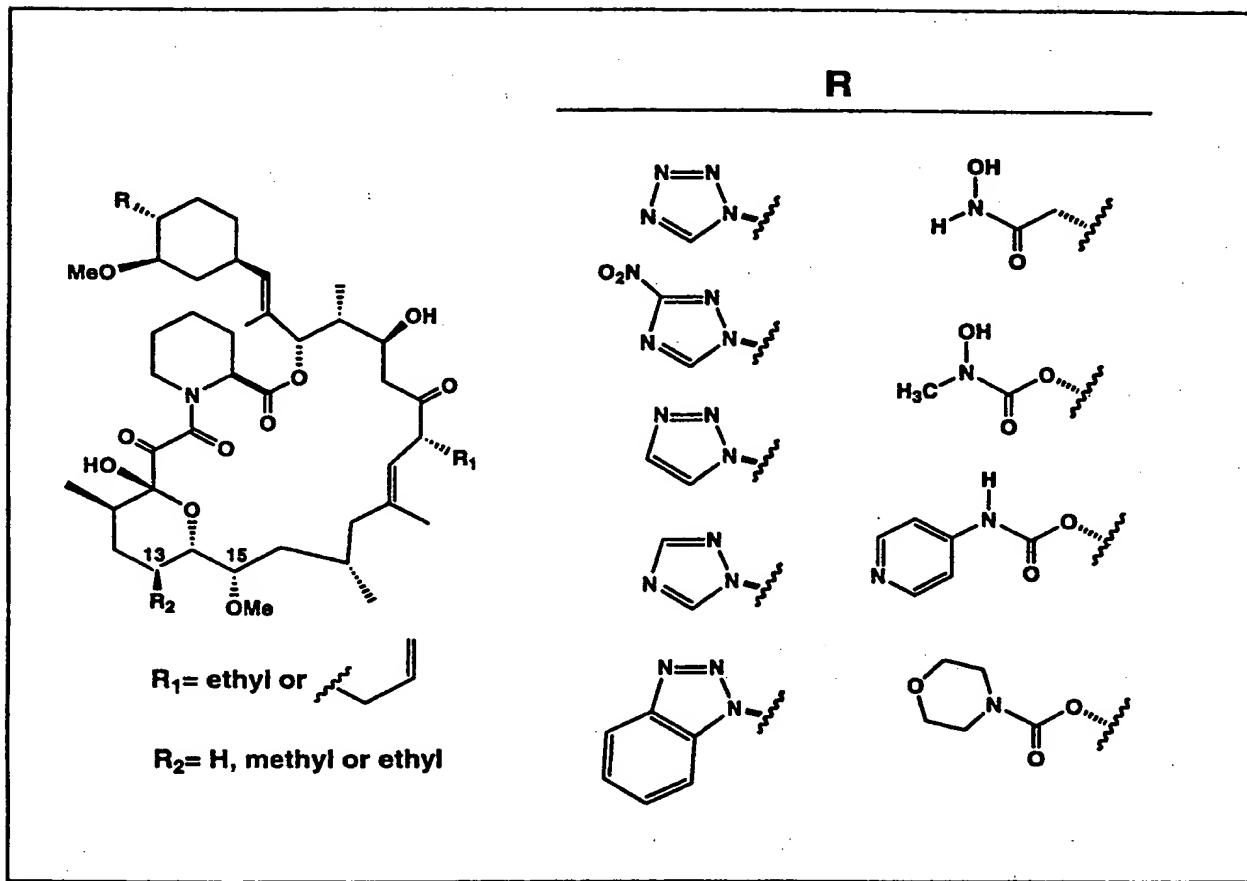


Figure 8
Part A

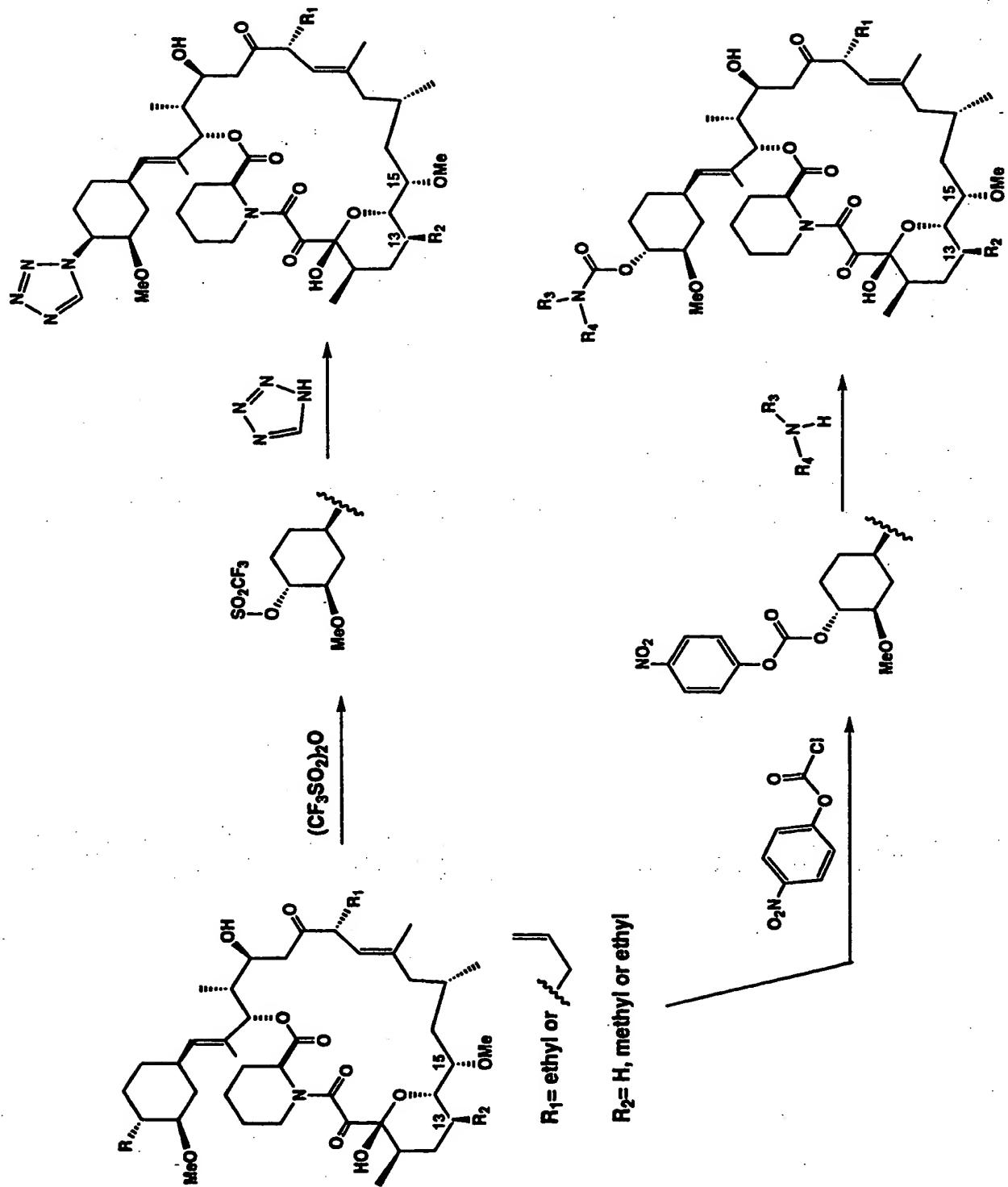


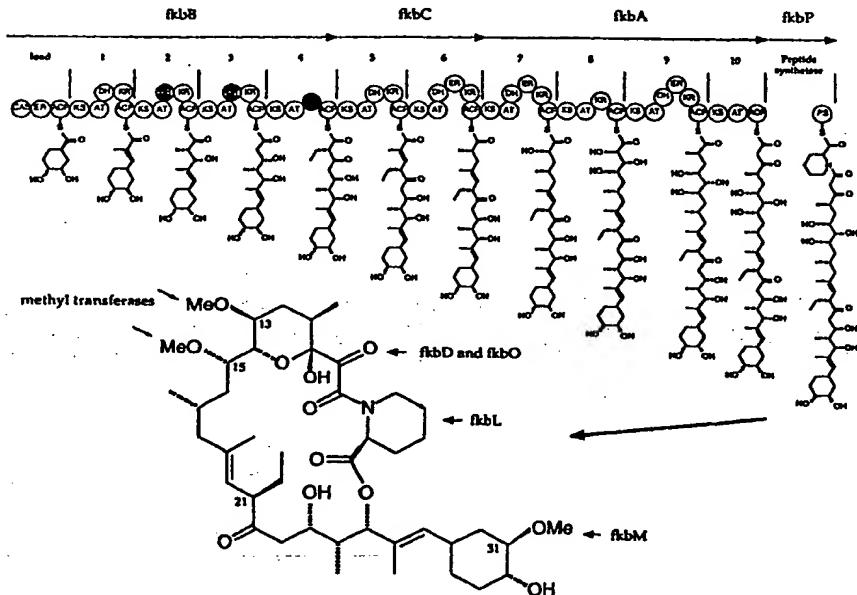
Figure 6
Part B



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ :		A2	(11) International Publication Number:	WO 00/20601
C12N 15/52, 15/54, 15/62, 9/10, C12P 17/18, 19/32, C07D 498/18 // (C07D 498/18, 311:00, 273:00, 211:00)			(43) International Publication Date:	13 April 2000 (13.04.00)
(21) International Application Number:	PCT/US99/22886		(74) Agents:	FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).
(22) International Filing Date:	1 October 1999 (01.10.99)		(81) Designated States:	AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, IL, IS, JP, KG, KP, KR, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(30) Priority Data:			Published	
60/102,748 60/123,810 60/139,650	2 October 1998 (02.10.98) 11 March 1999 (11.03.99) 17 June 1999 (17.06.99)	US US US	Without international search report and to be republished upon receipt of that report. With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.	
(71) Applicant (for all designated States except US):	KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).			
(72) Inventors; and				
(75) Inventors/Applicants (for US only):	REEVES, Christopher [US/US]; 4 East Altarinda Drive, Orinda, CA 94563 (US). CHU, Daniel [US/US]; 3767 Benton Street, Santa Clara, CA 95051 (US). KHOSLA, Chaitan [IN/US]; 740 Para Avenue, Palo Alto, CA 94306 (US). SANTI, Daniel [US/US]; 211 Belgrave Avenue, San Francisco, CA 94117 (US). WU, Kai [CN/US]; 900 Constitution Drive, Foster City, CA 94404 (US).			

(54) Title: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR



(57) Abstract

Host cells comprising recombinant vectors encoding the FK-520 polyketide synthase and FK-520 modification enzymes can be used to produce the FK-520 polyketide. Recombinant DNA constructs comprising one or more FK-520 polyketide synthase domains, modules, open reading frames, and variants thereof can be used to produce recombinant polyketide synthases and a variety of different polyketides with application as pharmaceutical and veterinary products.

*(Referred to in PCT Gazette No. 35/2000, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS
THEREFOR**

5

Field of the Invention

The present invention relates to polyketides and the polyketide synthase (PKS) enzymes that produce them. The invention also relates generally to genes encoding PKS enzymes and to recombinant host cells containing such genes and in which expression of such genes leads to the production of polyketides. The present invention also relates to 10 compounds useful as medicaments having immunosuppressive and/or neurotrophic activity. Thus, the invention relates to the fields of chemistry, molecular biology, and agricultural, medical, and veterinary technology.

Background of the Invention

15 Polyketides are a class of compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. Polyketides are biologically active molecules with a wide variety of structures, and the class encompasses numerous compounds with diverse activities. Tetracycline, erythromycin, 20 epothilone, FK-506, FK-520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of polyketides. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds.

25 This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce molecules that are structurally 30 related to, but distinct from, the polyketides produced from known PKS gene clusters. See, e.g., PCT publication Nos. WO 93/13663; 95/08548; 96/40968; 97/02358; 98/27203; and 98/49315; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and Fu *et al.*, 1994, *Biochemistry* 33:

9321-9326; McDaniel *et al.*, 1993, *Science* 262: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by PKS enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters, such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKS enzymes that incorporate such building blocks include an activity that functions as an amino acid ligase (an AMP ligase) or as a non-ribosomal peptide synthetase (NRPS). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesized. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

In the Type I or modular PKS enzyme group, a set of separate catalytic active sites (each active site is termed a "domain", and a set thereof is termed a "module") exists for each cycle of carbon chain elongation and modification in the polyketide synthesis pathway. The typical modular PKS is composed of several large polypeptides, which can be segregated from amino to carboxy termini into a loading module, multiple extender

modules, and a releasing (or thioesterase) domain. The PKS enzyme known as 6-deoxyerythronolide B synthase (DEBS) is a Type I PKS. In DEBS, there is a loading module, six extender modules, and a thioesterase (TE) domain. The loading module, six extender modules, and TE of DEBS are present on three separate proteins (designated DEBS-1, DEBS-2, and DEBS-3, with two extender modules per protein). Each of the DEBS polypeptides is encoded by a separate open reading frame (ORF) or gene; these genes are known as *eryAII*, *eryAII*, and *eryAIII*. See Caffrey *et al.*, 1992, *FEBS Letters* 304: 205, and U.S. Patent No. 5,824,513, each of which is incorporated herein by reference.

Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The loading module of DEBS consists of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. Another type of loading module utilizes an inactivated ketosynthase (KS) domain and AT and ACP domains. This inactivated KS is in some instances called KS^Q, where the superscript letter is the abbreviation for the amino acid, glutamine, that is

present instead of the active site cysteine required for ketosynthase activity. In other PKS enzymes, including the FK-506 PKS, the loading module incorporates an unusual starter unit and is composed of a CoA ligase like activity domain. In any event, the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl or 5 other acyl-CoA) and transfers it as a thiol ester to the ACP of the loading module.

The AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and 2-hydroxymalonyl) and transfers it to the ACP of that extender module to form a thioester. Each extender module is responsible for accepting a compound from a prior module, 10 binding a building block, attaching the building block to the compound from the prior module, optionally performing one or more additional functions, and transferring the resulting compound to the next module.

Each extender module of a modular PKS contains a KS, AT, ACP, and zero, one, two, or three domains that modify the beta-carbon of the growing polyketide chain. A 15 typical (non-loading) minimal Type I PKS extender module is exemplified by extender module three of DEBS, which contains a KS domain, an AT domain, and an ACP domain. These three domains are sufficient to activate a 2-carbon extender unit and attach it to the growing polyketide molecule. The next extender module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next 20 extender module until synthesis is complete.

Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module is transferred to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module one possesses an acyl-KS and a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then 25 covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading building block (elongation or extension).

The polyketide chain, growing by two carbons each extender module, is sequentially passed as covalently bound thiol esters from extender module to extender module, in an 30 assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly, however, additional enzymatic activities modify the beta

keto group of each two carbon unit just after it has been added to the growing polyketide chain but before it is transferred to the next module.

Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, and as noted above, other domains that modify the beta-carbonyl moiety can be present. Thus, modules may contain a ketoreductase (KR) domain that reduces the keto group to an alcohol. Modules may also contain a KR domain plus a dehydratase (DH) domain that dehydrates the alcohol to a double bond. Modules may also contain a KR domain, a DH domain, and an enoylreductase (ER) domain that converts the double bond product to a saturated single bond using the beta carbon as a methylene function. An extender module can also contain other enzymatic activities, such as, for example, a methylase or dimethylase activity.

After traversing the final extender module, the polyketide encounters a releasing domain that cleaves the polyketide from the PKS and typically cyclizes the polyketide. For example, final synthesis of 6-dEB is regulated by a TE domain located at the end of extender module six. In the synthesis of 6-dEB, the TE domain catalyzes cyclization of the macrolide ring by formation of an ester linkage. In FK-506, FK-520, rapamycin, and similar polyketides, the TE activity is replaced by a RapP (for rapamycin) or RapP like activity that makes a linkage incorporating a pipecolate acid residue. The enzymatic activity that catalyzes this incorporation for the rapamycin enzyme is known as RapP, encoded by the *rapP* gene. The polyketide can be modified further by tailoring enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated at C-6 and C-12 and glycosylated at C-3 and C-5 in the synthesis of erythromycin A.

In Type I PKS polypeptides, the order of catalytic domains is conserved. When all beta-keto processing domains are present in a module, the order of domains in that module from N-to-C-terminus is always KS, AT, DH, ER, KR, and ACP. Some or all of the beta-keto processing domains may be missing in particular modules, but the order of the domains present in a module remains the same. The order of domains within modules is believed to be important for proper folding of the PKS polypeptides into an active complex. Importantly, there is considerable flexibility in PKS enzymes, which allows for the genetic engineering of novel catalytic complexes. The engineering of these enzymes is achieved by modifying, adding, or deleting domains, or replacing them with those taken from other Type I PKS enzymes. It is also achieved by deleting, replacing, or adding entire modules with those

taken from other sources. A genetically engineered PKS complex should of course have the ability to catalyze the synthesis of the product predicted from the genetic alterations made.

Alignments of the many available amino acid sequences for Type I PKS enzymes has approximately defined the boundaries of the various catalytic domains. Sequence 5 alignments also have revealed linker regions between the catalytic domains and at the N- and C-termini of individual polypeptides. The sequences of these linker regions are less well conserved than are those for the catalytic domains, which is in part how linker regions are identified. Linker regions can be important for proper association between domains and between the individual polypeptides that comprise the PKS complex. One can thus view the 10 linkers and domains together as creating a scaffold on which the domains and modules are positioned in the correct orientation to be active. This organization and positioning, if retained, permits PKS domains of different or identical substrate specificities to be substituted (usually at the DNA level) between PKS enzymes by various available methodologies. In selecting the boundaries of, for example, an AT replacement, one can 15 thus make the replacement so as to retain the linkers of the recipient PKS or to replace them with the linkers of the donor PKS AT domain, or, preferably, make both constructs to ensure that the correct linker regions between the KS and AT domains have been included in at least one of the engineered enzymes. Thus, there is considerable flexibility in the 20 design of new PKS enzymes with the result that known polyketides can be produced more effectively, and novel polyketides useful as pharmaceuticals or for other purposes can be made.

By appropriate application of recombinant DNA technology, a wide variety of polyketides can be prepared in a variety of different host cells provided one has access to nucleic acid compounds that encode PKS proteins and polyketide modification enzymes. 25 The present invention helps meet the need for such nucleic acid compounds by providing recombinant vectors that encode the FK-520 PKS enzyme and various FK-520 modification enzymes. Moreover, while the FK-506 and FK-520 polyketides have many useful activities, there remains a need for compounds with similar useful activities but with better pharmacokinetic profile and metabolism and fewer side-effects. The present invention helps 30 meet the need for such compounds as well.

Summary of the Invention

In one embodiment, the present invention provides recombinant DNA vectors that encode all or part of the FK-520 PKS enzyme. Illustrative vectors of the invention include cosmid pKOS034-120, pKOS034-124, pKOS065-C31, pKOS065-C3, pKOS065-M27, and pKOS065-M21. The invention also provides nucleic acid compounds that encode the various domains of the FK-520 PKS, i.e., the KS, AT, ACP, KR, DH, and ER domains. These compounds can be readily used, alone or in combination with nucleic acids encoding other FK-520 or non-FK-520 PKS domains, as intermediates in the construction of recombinant vectors that encode all or part of PKS enzymes that make novel polyketides.

The invention also provides isolated nucleic acids that encode all or part of one or more modules of the FK-520 PKS, each module comprising a ketosynthase activity, an acyl transferase activity, and an acyl carrier protein activity. The invention provides an isolated nucleic acid that encodes one or more open reading frames of FK-520 PKS genes, said open reading frames comprising coding sequences for a CoA ligase activity, an NRPS activity, or two or more extender modules. The invention also provides recombinant expression vectors containing these nucleic acids.

In another embodiment, the invention provides isolated nucleic acids that encode all or a part of a PKS that contains at least one module in which at least one of the domains in the module is a domain from a non-FK-520 PKS and at least one domain is from the FK-520 PKS. The non-FK-520 PKS domain or module originates from the rapamycin PKS, the FK-506 PKS, DEBS, or another PKS. The invention also provides recombinant expression vectors containing these nucleic acids.

In another embodiment, the invention provides a method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector that encodes at least one module of a PKS, said module comprising at least one FK-520 PKS domain, and culturing said host cell under conditions such that said PKS is produced and catalyzes synthesis of said polyketide. In one aspect, the method is practiced with a *Streptomyces* host cell. In another aspect, the polyketide produced is FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-506 or rapamycin.

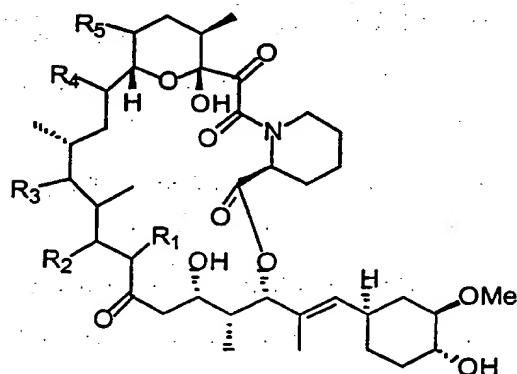
In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of ethylmalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require ethylmalonyl CoA for biosynthesis.

The invention also provides recombinant nucleic acids that encode AT domains specific for ethylmalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring ethylmalonyl CoA in host cells that otherwise are unable to produce such polyketides.

5 In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require 2-hydroxymalonyl CoA for biosynthesis. The invention also provides recombinant
10 nucleic acids that encode AT domains specific for 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring 2-hydroxymalonyl CoA or 2-methoxymalonyl CoA in host cells that are otherwise unable to produce such polyketides.

In another embodiment, the invention provides a compound related in structure to
15 FK-520 or FK-506 that is useful in the treatment of a medical condition. These compounds include compounds in which the C-13 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. Such compounds are less susceptible to the main *in vivo* pathway of degradation for FK-520 and FK-506 and related compounds and thus exhibit an improved pharmacokinetic profile. The compounds of the
20 invention also include compounds in which the C-15 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. The compounds of the invention also include the above compounds further modified by chemical methodology to produce derivatives such as, but not limited to, the C-18 hydroxyl derivatives, which have potent neurotrophin but not immunosuppression activities.

25 Thus, the invention provides polyketides having the structure:



wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506. The invention provides these compounds in purified form and in pharmaceutical compositions.

In another embodiment, the invention provides a method for treating a medical condition by administering a pharmaceutically efficacious dose of a compound of the invention. The compounds of the invention may be administered to achieve immunosuppression or to stimulate nerve growth and regeneration.

These and other embodiments and aspects of the invention will be more fully understood after consideration of the attached Drawings and their brief description below, together with the detailed description, examples, and claims that follow.

15

Brief Description of the Drawings

Figure 1 shows a diagram of the FK-520 biosynthetic gene cluster. The top line provides a scale in kilobase pairs (kb). The second line shows a restriction map with selected restriction enzyme recognition sequences indicated. K is *KpnI*; X is *XhoI*, S is *SacI*; P is *PstI*; and E is *EcoRI*. The third line indicates the position of FK-520 PKS and related genes. Genes are abbreviated with a one letter designation, i.e., C is *fkbC*. Immediately under the third line are numbered segments showing where the loading module (L) and ten different extender modules (numbered 1 - 10) are encoded on the various genes shown. At the bottom of the Figure, the DNA inserts of various cosmids of the invention (i.e., 34-124 is cosmid pKOS034-124) are shown in alignment with the FK-520 biosynthetic gene cluster.

Figure 2 shows the loading module (load), the ten extender modules, and the peptide synthetase domain of the FK-520 PKS, together with, on the top line, the genes that encode the various domains and modules. Also shown are the various intermediates in FK-520 biosynthesis, as well as the structure of FK-520, with carbons 13, 15, 21, and 31 numbered. The various domains of each module and subdomains of the loading module are also shown. The darkened circles showing the DH domains in modules 2, 3, and 4 indicate that the dehydratase domain is not functional as a dehydratase; this domain may affect the

stereochemistry at the corresponding position in the polyketide. The substituents on the FK-520 structure that result from the action of non-PKS enzymes are also indicated by arrows, together with the types of enzymes or the genes that code for the enzymes that mediate the action. Although the methyltransferase is shown acting at the C-13 and C-15 hydroxyl groups after release of the polyketide from the PKS, the methyltransferase may act on the 2-hydroxymalonyl substrate prior to or contemporaneously with its incorporation during polyketide synthesis.

Figure 3 shows a close-up view of the left end of the FK-520 gene cluster, which contains at least ten additional genes. The ethyl side chain on carbon 21 of FK-520 (Figure 2) is derived from an ethylmalonyl CoA extender unit that is incorporated by an ethylmalonyl specific AT domain in extender module 4 of the PKS. At least four of the genes in this region code for enzymes involved in ethylmalonyl biosynthesis. The polyhydroxybutyrate depolymerase is involved in maintaining hydroxybutyryl-CoA pools during FK-520 production. Polyhydroxybutyrate accumulates during vegetative growth and disappears during stationary phase in other *Streptomyces* (Ranade and Vining, 1993, *Can. J. Microbiol.* 39:377). Open reading frames with unknown function are indicated with a question mark.

Figure 4 shows a biosynthetic pathway for the biosynthesis of ethylmalonyl CoA from acetoacetyl CoA consistent with the function assigned to four of the genes in the FK-520 gene cluster shown in Figure 3.

Figure 5 shows a close-up view of the right-end of the FK-520 PKS gene cluster (and of the sequences on cosmid pKOS065-C31). The genes shown include *fkbD*, *fkbM* (a methyl transferase that methylates the hydroxyl group on C-31 of FK-520), *fkbN* (a homolog of a gene described as a regulator of cholesterol oxidase and that is believed to be a transcriptional activator), *fkbQ* (a type II thioesterase, which can increase polyketide production levels), and *fkbS* (a crotonyl-CoA reductase involved in the biosynthesis of ethylmalonyl CoA).

Figure 6 shows the proposed degradative pathway for tacrolimus (FK-506) metabolism.

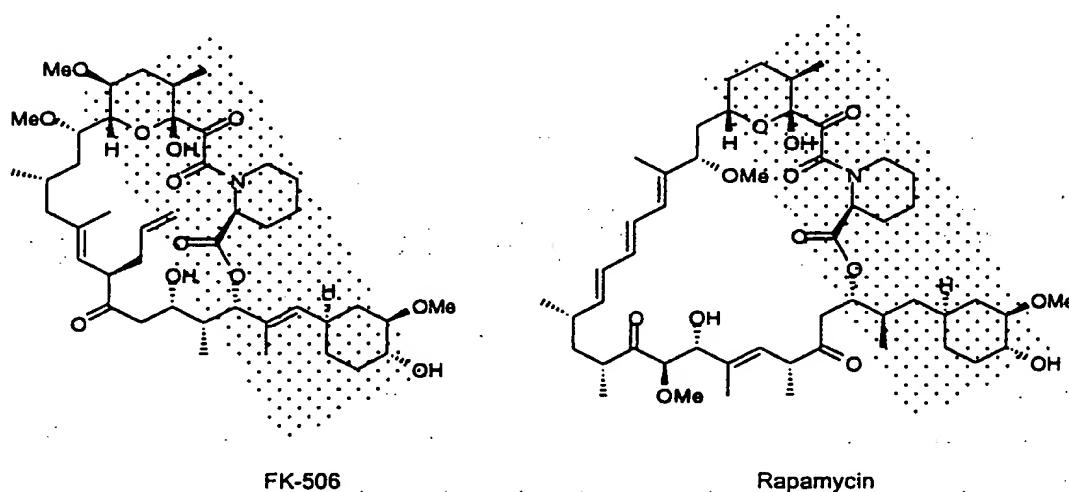
Figure 7 shows a schematic process for the construction of recombinant PKS genes of the invention that encode PKS enzymes that produce 13-desmethoxy FK-506 and FK-520 polyketides of the invention, as described in Example 4, below.

Figure 8, in Parts A and B, shows certain compounds of the invention preferred for dermal application in Part A and a synthetic route for making those compounds in Part B.

Detailed Description of the Invention

- 5 Given the valuable pharmaceutical properties of polyketides, there is a need for methods and reagents for producing large quantities of polyketides, as well as for producing related compounds not found in nature. The present invention provides such methods and reagents, with particular application to methods and reagents for producing the polyketides known as FK-520, also known as ascomycin or L-683,590 (see Holt *et al.*, 1993, *JACS* 115:9925), and FK-506, also known as tacrolimus. Tacrolimus is a macrolide immunosuppressant used to prevent or treat rejection of transplanted heart, kidney, liver, lung, pancreas, and small bowel allografts. The drug is also useful for the prevention and treatment of graft-versus-host disease in patients receiving bone marrow transplants, and for the treatment of severe, refractory uveitis. There have been additional reports of the 10 unapproved use of tacrolimus for other conditions, including alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, and scleroderma. The invention provides methods and reagents for making novel polyketides related in structure to FK-520 and FK-506, and structurally related polyketides such as rapamycin.

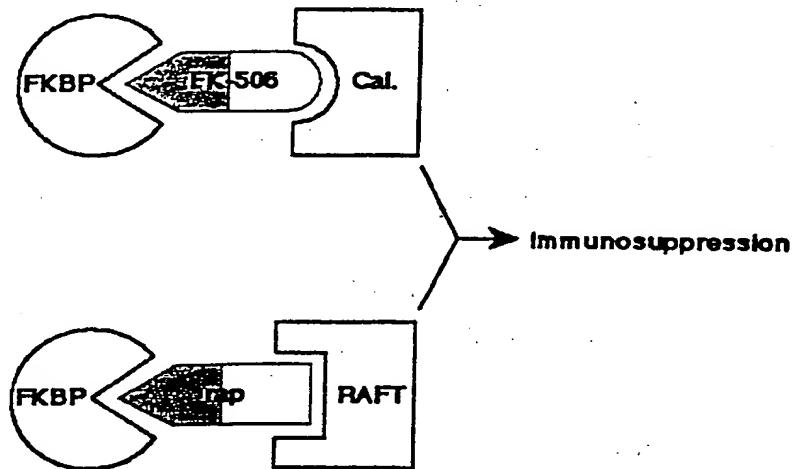
15 The FK-506 and rapamycin polyketides are potent immunosuppressants, with chemical structures shown below.



FK-520 differs from FK-506 in that it lacks the allyl group at C-21 of FK-506, having instead an ethyl group at that position, and has similar activity to FK-506, albeit reduced immunosuppressive activity.

These compounds act through initial formation of an intermediate complex with protein "immunophilins" known as FKBP (FK-506 binding proteins), including FKBP-12. Immunophilins are a class of cytosolic proteins that form complexes with molecules such as FK-506, FK-520, and rapamycin that in turn serve as ligands for other cellular targets involved in signal transduction. Binding of FK-506, FK-520, and rapamycin to FKBP occurs through the structurally similar segments of the polyketide molecules, known as the "FKBP-binding domain" (as generally but not precisely indicated by the stippled regions in the structures above). The FK-506-FKBP complex then binds calcineurin, while the rapamycin-FKBP complex binds to a protein known as RAFT-1. Binding of the FKBP-polyketide complex to these second proteins occurs through the dissimilar regions of the drugs known as the "effector" domains.

15



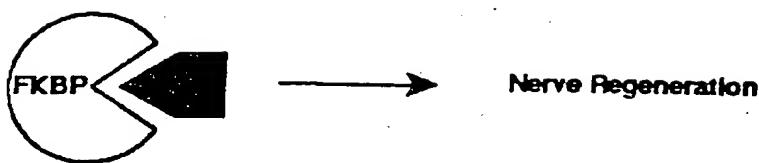
20

The three component FKBP-polyketide-effector complex is required for signal transduction and subsequent immunosuppressive activity of FK-506, FK-520, and rapamycin. Modifications in the effector domains of FK-506, FK-520, and rapamycin that destroy binding to the effector proteins (calcineurin or RAFT) lead to loss of immunosuppressive activity, even though FKBP binding is unaffected. Further, such analogs antagonize the immunosuppressive effects of the parent polyketides, because they compete for FKBP. Such non-immunosuppressive analogs also show reduced toxicity (see Dumont *et al.*, 1992, *Journal of Experimental Medicine* 176, 751-760), indicating that much of the toxicity of these drugs is not linked to FKBP binding.

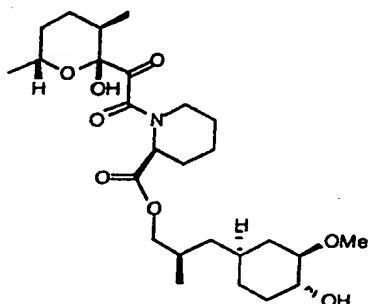
In addition to immunosuppressive activity, FK-520, FK-506, and rapamycin have neurotrophic activity. In the central nervous system and in peripheral nerves, immunophilins are referred to as "neuroimmunophilins". The neuroimmunophilin FKBP is markedly enriched in the central nervous system and in peripheral nerves. Molecules that bind to the 5 neuroimmunophilin FKBP, such as FK-506 and FK-520, have the remarkable effect of stimulating nerve growth. *In vitro*, they act as neurotrophins, i.e., they promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. See Gold *et al.*, Jun. 1999, *J. Pharm. Exp. Ther.* 289(3): 1202-1210; Lyons *et al.*, 1994, *Proc. National Academy of Science* 91: 3191-3195; Gold *et al.*, 1995, *Journal of Neuroscience* 15: 7509-7516; and Steiner *et al.*, 1997, *Proc. National Academy of Science* 94: 2019-2024. Further, the restored central and peripheral neurons appear to be functional.

Compared to protein neurotrophic molecules (BNDF, NGF, etc.), the small-molecule neurotrophins such as FK-506, FK-520, and rapamycin have different, and often advantageous, properties. First, whereas protein neurotrophins are difficult to deliver to their intended site of action and may require intra-cranial injection, the small-molecule neurotrophins display excellent bioavailability; they are active when administered subcutaneously and orally. Second, whereas protein neurotrophins show quite specific effects, the small-molecule neurotrophins show rather broad effects. Finally, whereas protein neurotrophins often show effects on normal sensory nerves, the small-molecule neurotrophins do not induce aberrant sprouting of normal neuronal processes and seem to affect damaged nerves specifically. Neuroimmunophilin ligands have potential therapeutic utility in a variety of disorders involving nerve degeneration (e.g. multiple sclerosis, 20 Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, peripheral neuropathies).

Recent studies have shown that the immunosuppressive and neurite outgrowth activity of FK-506, FK-520, and rapamycin can be separated; the neuroregenerative activity in the absence of immunosuppressive activity is retained by agents which bind to FKBP but not to the effector proteins calcineurin or RAFT. See Steiner *et al.*, 1997, *Nature Medicine* 30 3: 421-428.



Available structure-activity data show that the important features for neurotrophic activity of rapamycin, FK-520, and FK-506 lie within the common, contiguous segments of the macrolide ring that bind to FKBP. This portion of the molecule is termed the "FKBP binding domain" (see VanDuyne *et al.*, 1993, *Journal of Molecular Biology* 229: 105-124.). Nevertheless, the effector domains of the parent macrolides contribute to conformational rigidity of the binding domain and thus indirectly contribute to FKBP binding.

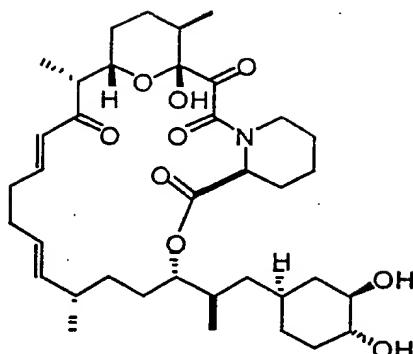


"FKBP binding domain"

There are a number of other reported analogs of FK-506, FK-520, and rapamycin that bind
10 to FKBP but not the effector protein calcineurin or RAPT. These analogs show effects on
nerve regeneration without immunosuppressive effects.

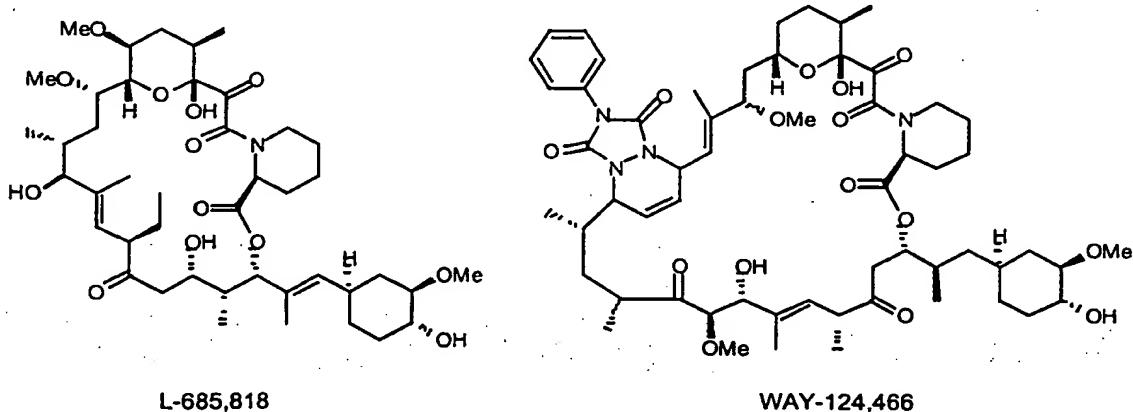
Naturally occurring FK-520 and FK-506 analogs include the antascomycins, which
are FK-506-like macrolides that lack the functional groups of FK-506 that bind to
calcineurin (see Fehr *et al.*, 1996, *The Journal of Antibiotics* 49: 230-233). These molecules
15 bind FKBP as effectively as does FK-506; they antagonize the effects of both FK-506 and
rapamycin, yet lack immunosuppressive activity.

14



Antascomycin A

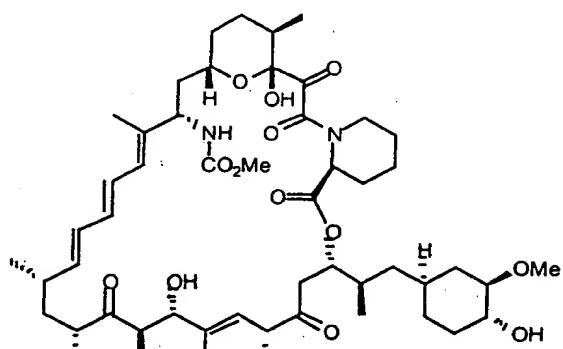
Other analogs can be produced by chemically modifying FK-506, FK-520, or rapamycin. One approach to obtaining neuroimmunophilin ligands is to destroy the effector binding region of FK-506, FK-520, or rapamycin by chemical modification. While the chemical modifications permitted on the parent compounds are quite limited, some useful chemically modified analogs exist. The FK-520 analog L-685,818 ($ED_{50} = 0.7 \text{ nM}$ for FKBP binding; see Dumont *et al.*, 1992), and the rapamycin analog WAY-124,466 ($IC_{50} = 12.5 \text{ nM}$; see Ocain *et al.*, 1993, *Biochemistry Biophysical Research Communications* 192: 1340-134693) are about as effective as FK-506, FK-520, and rapamycin at promoting neurite outgrowth in sensory neurons (see Steiner *et al.*, 1997).



One of the few positions of rapamycin that is readily amenable to chemical

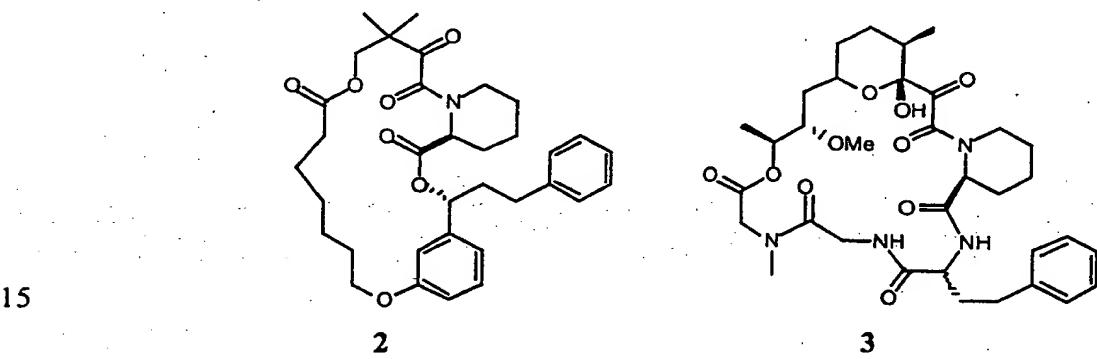
modification is the allylic 16-methoxy group; this reactive group is readily exchanged by acid-catalyzed nucleophilic substitution. Replacement of the 16-methoxy group of 15 rapamycin with a variety of bulky groups has produced analogs showing selective loss of immunosuppressive activity while retaining FKBP-binding (see Luengo *et al.*, 1995, *Chemistry & Biology* 2: 471-481). One of the best compounds, 1, below, shows complete

loss of activity in the splenocyte proliferation assay with only a 10-fold reduction in binding to FKBP.



1

5 There are also synthetic analogs of FKBP binding domains. These compounds reflect an approach to obtaining neuroimmunophilin ligands based on "rationally designed" molecules that retain the FKBP-binding region in an appropriate conformation for binding to FKBP, but do not possess the effector binding regions. In one example, the ends of the FKBP binding domain were tethered by hydrocarbon chains (see Holt *et al.*, 1993, *Journal of the American Chemical Society* 115: 9925-9938); the best analog, 2, below, binds to FKBP about as well as FK-506. In a similar approach, the ends of the FKBP binding domain were tethered by a tripeptide to give analog 3, below, which binds to FKBP about 20-fold poorer than FK-506. These compounds are anticipated to have neuroimmunophilin binding activity.



2

3

15 In a primate MPTP model of Parkinson's disease, administration of FKBP ligand GPI-1046 caused brain cells to regenerate and behavioral measures to improve. MPTP is a neurotoxin, which, when administered to animals, selectively damages nigral-striatal dopamine neurons in the brain, mimicking the damage caused by Parkinson's disease. Whereas, before treatment, animals were unable to use affected limbs, the FKBP ligand

restored the ability of animals to feed themselves and gave improvements in measures of locomotor activity, neurological outcome, and fine motor control. There were also corresponding increases in regrowth of damaged nerve terminals. These results demonstrate the utility of FKBP ligands for treatment of diseases of the CNS.

From the above description, two general approaches towards the design of non-immunosuppressant, neuroimmunophilin ligands can be seen. The first involves the construction of constrained cyclic analogs of FK-506 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. The advantages of this approach are that the conformation of the analogs can be accurately modeled and predicted by computational methods, and the analogs closely resemble parent molecules that have proven pharmacological properties. A disadvantage is that the difficult chemistry limits the numbers and types of compounds that can be prepared. The second approach involves the trial and error construction of acyclic analogs of the FKBP binding domain by conventional medicinal chemistry. The advantages to this approach are that the chemistry is suitable for production of the numerous compounds needed for such interactive chemistry-bioassay approaches. The disadvantages are that the molecular types of compounds that have emerged have no known history of appropriate pharmacological properties, have rather labile ester functional groups, and are too conformationally mobile to allow accurate prediction of conformational properties.

The present invention provides useful methods and reagents related to the first approach, but with significant advantages. The invention provides recombinant PKS genes that produce a wide variety of polyketides that cannot otherwise be readily synthesized by chemical methodology alone. Moreover, the present invention provides polyketides that have either or both of the desired immunosuppressive and neurotrophic activities, some of which are produced only by fermentation and others of which are produced by fermentation and chemical modification. Thus, in one aspect, the invention provides compounds that optimally bind to FKBP but do not bind to the effector proteins. The methods and reagents of the invention can be used to prepare numerous constrained cyclic analogs of FK-520 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. Such compounds will show neuroimmunophilin binding (neurotrophic) but not immunosuppressive effects. The invention also allows direct manipulation of FK-520 and related chemical structures *via* genetic engineering of the enzymes involved in the biosynthesis of FK-520 (as well as related compounds, such as FK-506 and rapamycin);

similar chemical modifications are simply not possible because of the complexity of the structures. The invention can also be used to introduce "chemical handles" into normally inert positions that permit subsequent chemical modifications.

Several general approaches to achieve the development of novel neuroimmunophilin
5 ligands are facilitated by the methods and reagents of the present invention. One approach is to make "point mutations" of the functional groups of the parent FK-520 structure that bind to the effector molecules to eliminate their binding potential. These types of structural modifications are difficult to perform by chemical modification, but can be readily accomplished with the methods and reagents of the invention.

10 A second, more extensive approach facilitated by the present invention is to utilize molecular modeling to predict optimal structures *ab initio* that bind to FKBP but not effector molecules. Using the available X-ray crystal structure of FK-520 (or FK-506) bound to FKBP, molecular modeling can be used to predict polyketides that should optimally bind to FKBP but not calcineurin. Various macrolide structures can be generated
15 by linking the ends of the FKBP-binding domain with "all possible" polyketide chains of variable length and substitution patterns that can be prepared by genetic manipulation of the FK-520 or FK-506 PKS gene cluster in accordance with the methods of the invention. The ground state conformations of the virtual library can be determined, and compounds that possess binding domains most likely to bind well to FKBP can be prepared and tested.

20 Once a compound is identified in accordance with the above approaches, the invention can be used to generate a focused library of analogs around the lead candidate, to "fine tune" the compound for optimal properties. Finally, the genetic engineering methods of the invention can be directed towards producing "chemical handles" that enable medicinal chemists to modify positions of the molecule previously inert to chemical
25 modification. This opens the path to previously prohibited chemical optimization of lead compounds by time-proven approaches.

Moreover, the present invention provides polyketide compounds and the recombinant genes for the PKS enzymes that produce the compounds that have significant advantages over FK-506 and FK-520 and their analogs. The metabolism and
30 pharmacokinetics of tacrolimus has been extensively studied, and FK-520 is believed to be similar in these respects. Absorption of tacrolimus is rapid, variable, and incomplete from the gastrointestinal tract (Harrison's Principles of Internal Medicine, 14th edition, 1998, McGraw Hill, 14, 20, 21, 64-67). The mean bioavailability of the oral dosage form is 27%,

(range 5 to 65%). The volume of distribution (VolD) based on plasma is 5 to 65 L per kg of body weight (L/kg), and is much higher than the VolD based on whole blood concentrations, the difference reflecting the binding of tacrolimus to red blood cells. Whole blood concentrations may be 12 to 67 times the plasma concentrations. Protein binding is 5 high (75 to 99%), primarily to albumin and alpha₁-acid glycoprotein. The half-life for distribution is 0.9 hour; elimination is biphasic and variable: terminal-11.3 hr (range, 3.5 to 40.5 hours). The time to peak concentration is 0.5 to 4 hours after oral administration.

Tacrolimus is metabolized primarily by cytochrome P450 3A enzymes in the liver and small intestine. The drug is extensively metabolized with less than 1% excreted 10 unchanged in urine. Because hepatic dysfunction decreases clearance of tacrolimus, doses have to be reduced substantially in primary graft non-function, especially in children. In addition, drugs that induce the cytochrome P450 3A enzymes reduce tacrolimus levels, while drugs that inhibit these P450s increase tacrolimus levels. Tacrolimus bioavailability doubles with co-administration of ketoconazole, a drug that inhibits P450 3A. See, Vincent 15 et al., 1992, *In vitro* metabolism of FK-506 in rat, rabbit, and human liver microsomes: Identification of a major metabolite and of cytochrome P450 3A as the major enzymes responsible for its metabolism, *Arch. Biochem. Biophys.* 294: 454-460; Iwasaki et al., 1993, Isolation, identification, and biological activities of oxidative metabolites of FK-506, a potent immunosuppressive macrolide lactone, *Drug Metabolism & Disposition* 21: 971-977; 20 Shiraga et al., 1994, Metabolism of FK-506, a potent immunosuppressive agent, by cytochrome P450 3A enzymes in rat, dog, and human liver microsomes, *Biochem. Pharmacol.* 47: 727-735; and Iwasaki et al., 1995, Further metabolism of FK-506 (Tacrolimus); Identification and biological activities of the metabolites oxidized at multiple sites of FK-506, *Drug Metabolism & Disposition* 23: 28-34. The cytochrome P450 3A 25 subfamily of isozymes has been implicated as important in this degradative process.

Structures of the eight isolated metabolites formed by liver microsomes are shown in Figure 6. Four metabolites of FK-506 involve demethylation of the oxygens on carbons 13, 15, and 31, and hydroxylation of carbon 12. The 13-demethylated (hydroxy) compounds undergo cyclizations of the 13-hydroxy at C-10 to give MI, MVI and MVII, and the 12-hydroxy metabolite at C-10 to give I. Another four metabolites formed by oxidation of the 30 four metabolites mentioned above were isolated by liver microsomes from dexamethasone treated rats. Three of these are metabolites doubly demethylated at the methoxy groups on carbons 15 and 31 (M-V), 13 and 31 (M-VI), and 13 and 15 (M-VII). The fourth, M-VIII,

was the metabolite produced after demethylation of the 31-methoxy group, followed by formation of a fused ring system by further oxidation. Among the eight metabolites, M-II has immunosuppressive activity comparable to that of FK-506, whereas the other metabolites exhibit weak or negligible activities. Importantly, the major metabolite of 5 human, dog, and rat liver microsomes is the 13-demethylated and cyclized FK-506 (M-I).

Thus, the major metabolism of FK-506 proceeds via 13-demethylation followed by cyclization to the inactive M-I, this representing about 90% of the metabolic products after a 10 minute incubation with liver microsomes. Analogs of tacrolimus that do not possess a C-13 methoxy group would not be susceptible to the first and most important 10 biotransformation in the destructive metabolism of tacrolimus (i.e. cyclization of 13-hydroxy to C-10). Thus, a 13-desmethoxy analog of FK-506 should have a longer half-life in the body than does FK-506. The C-13 methoxy group is believed not to be required for binding to FKBP or calcineurin. The C-13 methoxy is not present on the identical position 15 of rapamycin, which binds to FKBP with equipotent affinity as tacrolimus. Also, analysis of the 3-dimensional structure of the FKBP-tacrolimus-calcineurin complex shows that the C-13 methoxy has no interaction with FKBP and only a minor interaction with calcineurin. The present invention provides C-13-desmethoxy analogs of FK-506 and FK-520, as well 20 as the recombinant genes that encode the PKS enzymes that catalyze their synthesis and host cells that produce the compounds.

These compounds exhibit, relative to their naturally occurring counterparts, 25 prolonged immunosuppressive action *in vivo*, thereby allowing a lower dosage and/or reduced frequency of administration. Dosing is more predictable, because the variability in FK-506 dosage is largely due to variation of metabolism rate. FK-506 levels in blood can vary widely depending on interactions with drugs that induce or inhibit cytochrome P450 3A (summarized in USP Drug Information for the Health Care Professional). Of particular importance are the numerous drugs that inhibit or compete for CYP 3A, because they increase FK-506 blood levels and lead to toxicity (Prograf package insert, Fujisawa[®]US, Rev 4/97, Rec 6/97). Also important are the drugs that induce P450 3A (e.g. Dexamethasone), because they decrease FK-506 blood levels and reduce efficacy. Because 30 the major site of CYP 3A action on FK-506 is removed in the analogs provided by the present invention, those analogs are not as susceptible to drug interactions as the naturally occurring compounds.

Hyperglycemia, nephrotoxicity, and neurotoxicity are the most significant adverse effects resulting from the use of FK-506 and are believed to be similar for FK-520. Because these effects appear to occur primarily by the same mechanism as the immunosuppressive action (i.e. FKBP-calcineurin interaction), the intrinsic toxicity of the desmethoxy analogs 5 may be similar to FK-506. However, toxicity of FK-506 is dose related and correlates with high blood levels of the drug (Prograf package insert, Fujisawa US, Rev 4/97, Rec 6/97). Because the levels of the compounds provided by the present invention should be more controllable, the incidence of toxicity should be significantly decreased with the 13-desmethoxy analogs. Some reports show that certain FK-506 metabolites are more toxic 10 than FK-506 itself, and this provides an additional reason to expect that a CYP 3A resistant analog can have lower toxicity and a higher therapeutic index.

Thus, the present invention provides novel compounds related in structure to FK-506 and FK-520 but with improved properties. The invention also provides methods for making these compounds by fermentation of recombinant host cells, as well as the 15 recombinant host cells, the recombinant vectors in those host cells, and the recombinant proteins encoded by those vectors. The present invention also provides other valuable materials useful in the construction of these recombinant vectors that have many other important applications as well. In particular, the present invention provides the FK-520 PKS genes, as well as certain genes involved in the biosynthesis of FK-520 in recombinant form.

20 FK-520 is produced at relatively low levels in the naturally occurring cells, *Streptomyces hygroscopicus* var. *ascomyceticus*, in which it was first identified. Thus, another benefit provided by the recombinant FK-520 PKS and related genes of the present invention is the ability to produce FK-520 in greater quantities in the recombinant host cells provided by the invention. The invention also provides methods for making novel FK-520 25 analogs, in addition to the desmethoxy analogs described above, and derivatives in recombinant host cells of any origin.

The biosynthesis of FK-520 involves the action of several enzymes. The FK-520 PKS enzyme, which is composed of the *fkbA*, *fkbB*, *fkbC*, and *fkbP* gene products, 30 synthesizes the core structure of the molecule. There is also a hydroxylation at C-9 mediated by the P450 hydroxylase that is the *fkbD* gene product and that is oxidized by the *fkbO* gene product to result in the formation of a keto group at C-9. There is also a methylation at C-31 that is mediated by an O-methyltransferase that is the *fkbM* gene product. There are also methylations at the C-13 and C-15 positions by a methyltransferase believed to be encoded

by the *fkbG* gene; this methyltransferase may act on the hydroxymalonyl CoA substrates prior to binding of the substrate to the AT domains of the PKS during polyketide synthesis. The present invention provides the genes encoding these enzymes in recombinant form. The invention also provides the genes encoding the enzymes involved in ethylmalonyl CoA and 5 2-hydroxymalonyl CoA biosynthesis in recombinant form. Moreover, the invention provides *Streptomyces hygroscopicus* var. *ascomyceticus* recombinant host cells lacking one or more of these genes that are useful in the production of useful compounds.

The cells are useful in production in a variety of ways. First, certain cells make a useful FK-520-related compound merely as a result of inactivation of one or more of the 10 FK-520 biosynthesis genes. Thus, by inactivating the C-31 O-methyltransferase gene in *Streptomyces hygroscopicus* var. *ascomyceticus*, one creates a host cell that makes a desmethyl (at C-31) derivative of FK-520. Second, other cells of the invention are unable to make FK-520 or FK-520 related compounds due to an inactivation of one or more of the 15 PKS genes. These cells are useful in the production of other polyketides produced by PKS enzymes that are encoded on recombinant expression vectors and introduced into the host cell.

Moreover, if only one PKS gene is inactivated, the ability to produce FK-520 or an 20 FK-520 derivative compound is restored by introduction of a recombinant expression vector that contains the functional gene in a modified or unmodified form. The introduced gene produces a gene product that, together with the other endogenous and functional gene 25 products, produces the desired compound. This methodology enables one to produce FK-520 derivative compounds without requiring that all of the genes for the PKS enzyme be present on one or more expression vectors. Additional applications and benefits of such cells and methodology will be readily apparent to those of skill in the art after consideration of how the recombinant genes were isolated and employed in the construction of the compounds of the invention.

The FK-520 biosynthetic genes were isolated by the following procedure. Genomic DNA was isolated from *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) using the lysozyme/proteinase K protocol described in Genetic Manipulation of 30 *Streptomyces* - A Laboratory Manual (Hopwood *et al.*, 1986). The average size of the DNA was estimated to be between 80 - 120 kb by electrophoresis on 0.3% agarose gels. A library was constructed in the SuperCosTM vector according to the manufacturer's instructions and with the reagents provided in the commercially available kit (Stratagene). Briefly, 100 µg of

genomic DNA was partially digested with 4 units of *Sau3A* I for 20 min. in a reaction volume of 1 mL, and the fragments were dephosphorylated and ligated to SuperCos vector arms. The ligated DNA was packaged and used to infect log-stage XL1-BlueMR cells. A library of about 10,000 independent cosmid clones was obtained.

5 Based on recently published sequence from the FK-506 cluster (Motamed and Shafiee, 1998, *Eur. J. Biochem.* 256: 528), a probe for the *fkbO* gene was isolated from ATCC 14891 using PCR with degenerate primers. With this probe, a cosmid designated pKOS034-124 was isolated from the library. With probes made from the ends of cosmid pKOS034-124, an additional cosmid designated pKOS034-120 was isolated. These cosmids
10 (pKOS034-124 and pKOS034-120) were shown to contain DNA inserts that overlap with one another. Initial sequence data from these two cosmids generated sequences similar to sequences from the FK-506 and rapamycin clusters, indicating that the inserts were from the FK-520 PKS gene cluster. Two *EcoRI* fragments were subcloned from cosmids pKOS034-124 and pKOS034-120. These subclones were used to prepare shotgun libraries by partial
15 digestion with *Sau3AI*, gel purification of fragments between 1.5 kb and 3 kb in size, and ligation into the pLitmus28 vector (New England Biolabs). These libraries were sequenced using dye terminators on a Beckmann CEQ2000 capillary electrophoresis sequencer, according to the manufacturer's protocols.

20 To obtain cosmids containing sequence on the left and right sides of the sequenced region described above, a new cosmid library of ATCC 14891 DNA was prepared essentially as described above. This new library was screened with a new *fkbM* probe isolated using DNA from ATCC 14891. A probe representing the *fkbP* gene at the end of cosmid pKOS034-124 was also used. Several additional cosmids to the right of the previously sequenced region were identified. Cosmids pKOS065-C31 and pKOS065-C3
25 were identified and then mapped with restriction enzymes. Initial sequences from these cosmids were consistent with the expected organization of the cluster in this region. More extensive sequencing showed that both cosmids contained in addition to the desired sequences, other sequences not contiguous to the desired sequences on the host cell chromosomal DNA. Probing of additional cosmid libraries identified two additional
30 cosmids, pKOS065-M27 and pKOS065-M21, that contained the desired sequences in a contiguous segment of chromosomal DNA. Cosmids pKOS034-124, pKOS034-120, pKOS065-M27, and pKOS065-M21 have been deposited with the American Type Culture Collection, Manassas, VA, USA. The complete nucleotide sequence of the coding

sequences of the genes that encode the proteins of the FK-520 PKS are shown below but can also be determined from the cosmids of the invention deposited with the ATCC using standard methodology.

Referring to Figures 1 and 3, the FK-520 PKS gene cluster is composed of four open reading frames designated *fkbB*, *fkbC*, *fkbA*, and *fkbP*. The *fkbB* open reading frame encodes the loading module and the first four extender modules of the PKS. The *fkbC* open reading frame encodes extender modules five and six of the PKS. The *fkbA* open reading frame encodes extender modules seven, eight, nine, and ten of the PKS. The *fkbP* open reading frame encodes the NRPS of the PKS. Each of these genes can be isolated from the cosmids of the invention described above. The DNA sequences of these genes are provided below preceded by the following table identifying the start and stop codons of the open reading frames of each gene and the modules and domains contained therein.

	<u>Nucleotides</u>	<u>Gene or Domain</u>
15	complement (412 - 1836)	<i>fkbW</i>
	complement (2020 - 3579)	<i>fkbV</i>
	complement (3969 - 4496)	<i>fkbR2</i>
	complement (4595 - 5488)	<i>fkbR1</i>
	5601 - 6818	<i>fkbE</i>
20	6808 - 8052	<i>fkbF</i>
	8156 - 8824	<i>fkbG</i>
	complement (9122 - 9883)	<i>fkbH</i>
	complement (9894 - 10994)	<i>fkbI</i>
	complement (10987 - 11247)	<i>fkbJ</i>
25	complement (11244 - 12092)	<i>fkbK</i>
	complement (12113 - 13150)	<i>fkbL</i>
	complement (13212 - 23988)	<i>fkbC</i>
	complement (23992 - 46573)	<i>fkbB</i>
	46754 - 47788	<i>fkbO</i>
30	47785 - 52272	<i>fkbP</i>
	52275 - 71465	<i>fkbA</i>
	71462 - 72628	<i>fkbD</i>
	72625 - 73407	<i>fkbM</i>
	complement (73460 - 76202)	<i>fkbN</i>
35	complement (76336 - 77080)	<i>fkbQ</i>
	complement (77076 - 77535)	<i>fkbS</i>
	complement (44974 - 46573)	CoA ligase of loading domain
	complement (43777 - 44629)	ER of loading domain
	complement (43144 - 43660)	ACP of loading domain
40	complement (41842 - 43093)	KS of extender module 1 (KS1)
	complement(40609 - 41842)	AT1
	complement (39442 - 40609)	DH1
	complement (38677 - 39307)	KR1
	complement (38371 - 38581)	ACP1

	complement (37145 - 38296)	KS2
	complement (35749 - 37144)	AT2
	complement (34606 - 35749)	DH2 (inactive)
	complement (33823 - 34480)	KR2
5	complement (33505 - 33715)	ACP2
	complement (32185 - 33439)	KS3
	complement (31018 - 32185)	AT3
	complement (29869 - 31018)	DH3 (inactive)
	complement (29092 - 29740)	KR3
10	complement (28750 - 28960)	ACP3
	complement (27430 - 28684)	KS4
	complement (26146 - 27430)	AT4
	complement (24997 - 26146)	DH4 (inactive)
	complement (24163 - 24373)	ACP4
15	complement (22653 - 23892)	KS5
	complement (21420 - 22653)	AT5
	complement (20241 - 21420)	DH5
	complement (19464 - 20097)	KR5
	complement (19116 - 19326)	ACP5
20	complement (17820 - 19053)	KS6
	complement (16587 - 17820)	AT6
	complement (15438 - 16587)	DH6
	complement (14517 - 15294)	ER6
	complement (13761 - 14394)	KR6
25	complement (13452 - 13662)	ACP6
	52362 - 53576	KS7
	53577 - 54716	AT7
	54717 - 55871	DH7
	56019 - 56819	ER7
30	56943 - 57575	KR7
	57710 - 57920	ACP7
	57990 - 59243	KS8
	59244 - 60398	AT8
	60399 - 61412	DH8 (inactive)
35	61548 - 62180	KR8
	62328 - 62537	ACP8
	62598 - 63854	KS9
	63855 - 65084	AT9
	65085 - 66254	DH9
40	66399 - 67175	ER9
	67299 - 67931	KR9
	68094 - 68303	ACP9
	68397 - 69653	KS10
	69654 - 70985	AT10
45	71064 - 71273	ACP10

1 GATCTCAGGC ATGAAGTCCT CCAGGGGAGG CGCCGAGGTG GTGAACACCT CGCCGCTGCT
 61 TGTACGGACC ACTTCAGTCA GCGCGATTG CGGAACCAAG TCATCCGGAA TAAAGGGCGG
 121 TTACAAGATC CTCACATTGC GCGACCGCCA GCATACGCTG AGTTGCCTCA GAGGCAAACC
 181 GAAAGGGCGC GGGCGTCCG CACCAGGGCG GAGTACGCGA CGAGAGTGGC GCACCCCGCGC

	241	ACCGTCACCT	CTCTCCCCG	CCGGCGGGAT	GCCC	GGCGT	ACACGGTTGG	GCTCTCC	TG
	301	ACGCTGAACA	CCC	CGCGG	GTGGC	GTGCG	GGACAC	CACGCC	TGC
	361	TACGGGGAGG	GCGTACGGCG	GCCG	TGGCT	GTGCT	GGCG	CGGGT	GCACGG
5	421	GAGACGGCAC	TCGGCGAGCA	GGG	ACGCC	TCG	GGG	GGGCC	GACCGTGTG
	481	GTTCGGGGC	GGGCGGTG	CGG	TGGTG	AGG	GGGCGGT	AGG	CTGAGCG
	541	GTGACACGGC	AGCAAAGGC	GG	AGTCG	GGG	AAAGGT	TCG	ACGAGGG
	601	CGTCCGTCC	TCGATGCG	AGT	AGTCG	GGG	CGCTG	CGT	CGTGTG
	661	GCGTACACGT	CGGAGCCC	GGG	CGAGCAGG	GCAG	AGCGT	GAG	AGTGCCT
	721	CAGCGGCTT	CCGATACGAC	CGG	TCAACGC	GAT	CGT	ACGG	GGACGCCGGA
10	781	GGAGCGGGTG	GGGTAGTC	AGT	CGGCATC	GCAG	CCC	GGG	CGCAATA
	841	CGGTGTG	CGG	GCTT	CC	CCC	ATCGAA	GGCG	GGCGCAATA
	901	CTCGC	TAGA	CTCC	AGT	CGT	GTAC	GGG	GGACATACGC
	961	GAACCCGGC	CGGAGCAGG	CGG	AGCAGG	GA	AGTGC	GGG	ATGGTGT
15	1021	GGTGGGGTAG	TCGCGCAGG	CGG	CCGGCAG	GAAGG	TGAAAG	AGG	TTGGGAC
	1081	CCACAGGGTG	CCTCCCAGT	CGACT	CC	GTC	TACAGC	T	CTCCAGCTG
	1141	CCAGCGCACG	AGGTAGCCG	CGT	TGGACAT	CCC	GGTGACC	AGG	GTGCGCT
	1201	GTGGTAGCGC	TGGGCGACCG	ACG	CCGGG	GGC	CCC	GGG	CTGCCGCTA
	1261	CCACTCGGC	ACGGCGT	CCG	GGCGG	GCC	CATC	ACGG	GGCGGTGTT
20	1321	GCCCTTGTG	GTGGCGCGT	AGG	CGTAACC	GC	GGGCGAGC	ACC	CAGCTGG
	1381	GTCGTTGGC	TACTGCT	GGT	TACCGGG	GGT	CCG	GGG	CGATGGCCC
	1441	GCGGTCGGG	AGCCGGATG	CGA	ACTGGG	GTC	GTG	GGG	TGGTGTG
	1501	GGTGGAGGTG	TCGGGAAAGT	AGC	CGTCGAT	CTG	GGATC	CCG	TGGGAGTGG
	1561	CAGGTTCTT	GGCGTCA	CG	TCCCAGTC	CG	CCGGG	TCG	TGGCCCG
	1621	TCCC	GTG	GTCAGCT	CCAGG	GGC	CCT	GTG	GGGACACG
25	1681	CAGCTGGGAC	AGACGGG	AGT	GACCGTC	CGG	GGC	GTG	GGGCGTGGC
	1741	CGGTGAGGG	AGCAGGAC	CGA	CTGCG	CAG	GGG	TGAGA	GGCGCGAGG
	1801	TCTCGGGG	CGTCC	ACAC	CGAGGG	AAC	CAT	GGAG	CGCTCCAGA
	1861	GATGACGGAC	TGGAGGCT	GTC	CGC	GTG	AGAC	ACAT	GGGATG
	1921	ACTGAGGCC	CTCAGAGG	GGC	CCCG	ATG	ACGGG	GGG	GCTCC
30	1981	GGCGGTG	CCG	GGGCC	CCG	GGG	CCC	GGG	GTGCG
	2041	GACGGTGAAG	TAGCCG	TG	GCGACT	CAAGG	TG	GTG	GGTACAG
	2101	GCCCATG	TGG	CCGAGC	CTT	GGT	GTA	GG	GGCGG
	2161	CGCCTGG	ACG	GCGT	AGT	GGC	GGG	GGG	GTCTG
35	2221	CGCGGTG	ACC	GCGCC	GAGA	CGG	GTG	CCC	GGCGACCGC
	2281	GTAGGTGTG	GAT	GTG	CCCC	CC	TG	GGT	GGGACGT
	2341	GGTGATCTG	GCAC	CGT	GGT	GG	AGT	CGG	TGGGTTCCA
	2401	GGTCAGGCT	ATGG	GGTGT	CGG	GGG	GGT	GGG	AGGCGGAG
	2461	CGAACCGGG	TGGAGG	CGG	ATCCG	GCC	GAAGA	TGCG	TGAGCTG
40	2521	ACAGATCG	TCCAG	GAAGT	AGG	CGG	GGT	GCT	CTCCGGT
	2581	GGGATCG	ACC	GGGG	TG	CCC	GAT	CC	GGCCACCGA
	2641	TCCG	TG	GGT	ACT	CTC	GG	GG	GGAGGTACG
	2701	GTCCGG	TGGG	ACACG	CGT	GC	GG	CC	TCGCGGTT
	2761	GC	GGG	CGCG	AGG	TG	GG	CC	GACGGG
	2821	CGACCACG	GGG	TAGCC	CAC	GG	GGG	GG	TGAGTC
45	2881	CCC	GGG	GTTC	ATG	CGC	TG	GTG	GGT
	2941	GGCGACG	ACC	GGG	CGC	GG	GAAGA	CG	ACGT
	3001	GGCACCG	GGG	ACAG	GG	GTG	GG	GG	TGAGG
	3061	GACGGTGT	GA	GGG	CAT	GC	GGG	GG	GGTTG
	3121	GCTG	CTG	GG	AAC	GGT	GG	GG	TGTC
50	3181	CA	CGAGC	AGG	CCAT	GG	TGAG	GG	GGCGT
	3241	CT	GGG	GT	GG	GG	GG	GG	GGG
	3301	CG	GGG	CCG	GG	GG	GG	GG	GGG
	3361	GGT	CA	GGT	GG	GG	GG	GG	GGG
55	3421	CG	GGG	CCG	AGC	GGG	GGG	GGG	GGG
	3481	CA	CCCC	CCG	CTC	GG	GGG	GGG	GGG
	3541	CA	CGG	GGT	AGG	AT	GGT	GG	GGG
	3601	GG	GGG	ACAC	GG	GGG	GGG	GGG	GGG
	3661	TA	GGG	GTG	GG	GG	GG	GG	GGG
	3721	TG	CG	CC	GG	GG	GG	GG	GGG
60	3781	AC	CC	GA	AC	GG	GG	GG	GGG

3841	ACGGACCAGGG	CGTCGGCGGA	CCGGGCGTCG	GCGGGCTGGG	CGGTATGGCG	GCCGAGGACG	
3901	CCAGCCGCGT	GGGGCGGCCG	CGCCCAAGTG	CAGTACGCCG	ACCGTGGCCG	GCGGGAGGGC	
3961	CGGACCGGTC	AGTGCAGTCC	CGCGGCCCTG	CGGGACCGCT	CGTCCCAGAC	GGGTTCCACC	
4021	GCGGCCAAC	GGGGTCCGTG	TCCCGGGCGG	TAGACCATCA	GTGTCGGCTC	GAAGGTGATG	
5	4081	ACGATGACAC	CGTCCTGGTT	GTAGCCGATG	GTGCGCACCG	TGATGATGCC	TACGTCAAGGT
	4141	CGGCTGGCGG	ACTCCCGGGT	GTTCAGGACC	TCGGACTGCG	AGTAGATGGT	GTCGCCCTCG
	4201	AAGACCGGGT	TCGGCAGCCT	GACCCGGTCC	CAGCCGAGGT	TGGCCATCAC	ATGCTGGGAG
	4261	ATGTCGGTGA	CGCTCTGCC	GGTGACCAGG	GCGAGGGTGA	AGGTGGAGTC	CACCAGCGGC
10	4321	TTGCCCCAGG	TGGTGCCTCGC	CGAGTAGTGG	CGGTCGAAGT	GCAGCGGC	GGTGTCTGC
	4381	GTCAGGAGCG	TGAGCCAGGA	GTTGCGGTG	TCCAGGACCG	TGCGGCCAG	GGGGTGGCGG
	4441	TACACGTCGC	CGGTGGTGA	GTCCTCGAAG	TAGCGGCCCT	GCCAGCCCTC	GACCACAGCG
	4501	GTGCGGGTGG	CGTCCTGGTC	CGGGTTCTCA	GTGTCATGG	CGCTCATTCT	GGGAAGTCCC
15	4561	CGGTCCGCTG	TGAAATGCCG	AACCTTCACC	GGGTCATAC	GTGCGGC	TGAGCCCTGG
	4621	ACCGTACGTA	GTCGTAGAAC	CTCGCCACCA	CTGGCGCG	TGGTCCTCCG	GCGAGTGTGA
	4681	CCACGCCGAC	CGTGCCTCGC	GCCTGCGGGT	CGTCGAGCG	CACGGCGACG	GCGTGGTCAC
	4741	CGGGCCCGGA	CGGGCTGCCG	GTGAGGGGGG	CGACGGCCAC	ACCGAGGCCG	GCGGCGACCA
20	4801	GGGCCCCGAG	CGTGCCTAGC	TCGGTGCTCT	CCAGGACGAC	CCGCGGCACG	AATCCGGCG
	4861	CGGCGCACAG	CCGGTCGGTG	ATCTGGCGCA	GTCCGAAGAC	CGGCTCCAGT	GCCACGAACG
	4921	CCTCATCGGC	CAGCTCCGCG	GTCCGCACCC	GGCGCGTCT	GGCCAGCCG	TGTCCGGGTG
25	4981	GGACGAGCAG	GCACAGTGCC	TCGTCCTCGA	GTGGTGTCCA	CTCCACATCG	TCCCCGGCGG
	5041	GTCGTGGGCT	GGTCAGCCCC	AGGTCCAGCC	TGCTGTTGCG	GACGTCGTG	ACCACGGCGT
	5101	CGGCGGC	GCCGCGCAGT	TCGAAGGTGG	TGCCGGGAGC	CAGCCGGCGG	TACCCGGCGA
	5161	GGAGGTCGGG	CACCAAGCCAG	GTGCCGTAGG	AGTGCAGGAA	ACCCAGTGCC	ACGGTGCCGG
30	5221	TGTCGGGTG	GATCAGGGCG	GTGATGCGCT	GTCGGCGCC	GGAGACCTCA	CTGATCGCGC
	5281	GCAGGGCGTG	GGCGCGGAAG	ACCTCGCCGT	ACTTGTGAG	CCGGAGCCGG	TTCTGGTGC
	5341	GTCGAACAG	CGGCACGCC	ACTCGTCGCT	CCAGCCGCC	GATGGCCCTG	GACAGGGTCG
	5401	GCTGGGAGAT	GTTGAGCCG	TCCGCGGTGA	TCGTCACCGT	CTCGTGTCTG	GCCAAGGCCG
	5461	TGAACCACTG	CAACTCCCGT	ATCTCCATGC	AGGGACTATA	CGTACCGGGC	ATGGTCTCTGG
35	5521	CGAGGTTTCG	TCATTTACA	CGGGCCGGG	GGCGGCCAC	AGTGAGTCCT	CACCAACCAG
	5581	GACCCATGG	GAGGGACCCC	ATGTCGAGC	CCGATCCTCG	CCCTGAACAG	GAACGCCCG
	5641	CGGGGCC	GTCCGGTCTG	CTCGTGGTT	CTTGAGGCA	GGCCGTCG	GCTCCGGTC
	5701	CCACCCGCCA	CTGGCGGAC	CTGGCGGCC	GTGTCATCAA	GATCGAACGC	CCCAGCAGCG
	5761	GGCACCTCGC	CGCGGCC	GACCCACGG	TGCGTGGCAT	GTCCAGCCAC	TTCGTC
40	5821	TGAACCGGGG	GAAGGAGAGC	GTCCAGCTC	ATGTCGCTC	GCCGGAGGGC	AACCGGCACC
	5881	TGCACGCC	GGTGGACCGG	GCCGATGTC	TGGTGCAGAA	TCTGGCACCC	GGCGCCCGG
	5941	GCCGCTGGC	ATCGGCCACC	AGGTCTCGC	GCGGAGCCAC	CGAGGCTGAT	CACCTCGGA
	6001	CATATCCGGC	TACGGCAGTA	CCGGCTGCTA	CCGCGGACCG	CAAGGGCTAC	GACCTCCTGG
	6061	TCCAGTGC	AGCGGGGCTG	GTCTCCATCA	CCGGCACCCC	CGAGACCCG	TCCAAGGTGG
45	6121	GCCTGTC	CGCGGACATC	TGTGCGGG	TGTACGCGTA	CTCCGGC	CTCACGGCCC
	6181	TGCTGAAGCG	GGCCCGCACC	GGCCGGGGCT	CCGAGTTGGA	GGTCTCGATG	CTCGAAGCCC
	6241	TCGGTGAATG	GATGGGATAC	CCGAGACT	ACACGCGCTA	CGCCGGCACC	GCTCCGGCCC
	6301	GCGCCGGC	CAGCCACGCG	ACGATGCC	CCTACGCC	GTTACCA	CGCGACGGG
	6361	AGACGATCAA	TCTCGGGCTC	CAGAACGAGC	GGGAGTGGG	TTCTTCTG	GGTGTGGTGC
50	6421	TACAACGCC	CGGTCTCTGC	GACGACCCG	GCTTTCCGG	CAACGCCGAC	CGGGTGGCGC
	6481	ACCGCACCGA	GCTCGACGCC	CTGGTGA	AGGTGACGGG	CACGCTCACC	GGCGAGGAAC
	6541	TGGTGGCGC	GCTGGAGGAG	GCGTCGATG	CCTACGCA	CCAGCGCACC	GTGCGGGAGT
	6601	TCAGCGAAC	CCCCCAACTG	CGTACCG	GACGCTGGG	TCCGTTG	AGCCCGGTCG
	6661	GTGCGCTGG	GGGCCTGATC	CCCCCGTCA	CCTTCCACGG	CGAGCACCCG	GGCGGGCTGG
55	6721	GGCGGTCCC	GGAGCTGGG	GAGCATACCG	AGTCCGTCT	GGCGTGGCTG	GCCGCCCC
	6781	ACAGCGCCGA	CCCGAAGAG	GCCGCCATG	CCGAATGAAC	TCACCGGAGT	CCTGATCCTG
	6841	GCCGCGTGT	TCCCTGCTCG	CGCGTACGG	GGGCTGAACA	TGGGCCTG	CGCGCTGGTC
	6901	GCCACCTT	TGCTGGGGT	GGTCGCACTC	GACCGAACG	CGGACGAGGT	GCTGGGGG
	6961	TTCCCCCGA	GCATGTTCT	GGTGCTGGT	GGCGTCACGT	TCCTCTTC	GATCGCCCC
60	7021	GTCAACGGCA	CGGTGGACTG	GCTGGTACGT	GTGCGGGT	GGGCGGTGGG	GGCCCGGGGTG
	7081	GGAGCCGT	CCTGGGTGCT	CTTCGGCCTG	GGGGCACTG	TCTGCGC	AGGCGCCGCC
	7141	TCGCCCCG	CGGTGGCGAT	CGTGGCGCC	ATCAGCGT	CGTCG	CAGGCACCGC
	7201	ATCGATCCG	TGTCAGCCG	ACTGATGGG	GTGAACGGG	CCGCA	CAGTTCC
	7261	CCCTCCGGG	TCCCTGGCG	CATCGCCAC	TCGGCGT	AGAAGAACCA	TCTGCC
	7321	AGCGGGCGGG	TGCTCTTC	AGGCACCTC	GCCTTCA	TGGCGT	CGCGGTGTC
	7381	TGGCTCGT	TCGGCGCAG	GCGCCTCGA	CCACATGACC	TGGACGAGGA	CACCGATCCC

7441 ACGGAAGGGG ACCCGGCTTC CGGCCCGGC GCGGAACACG TGATGACGCT GACCGCGATG
 7501 GCGCGCTGG TGCTGGAAC CACGGTCCCTC TCCCTGGACA CCGGCTTCCT GGCCCTCAC
 7561 TTGGCGGCCT TGCTGGCGCT GCTCTTCCC CGCACCTCCC AGCAGGCCAC CAAGGAGATC
 7621 GCCTGGCCCG TGGTGTGCT GGTATGCGGG ATCGTGACCT ACGTCGCCCT GCTCCAGGAG
 5 7681 CTGGGCATCG TGGACTCCCT GGGGAAGATG ATCGCGGCCA TCGGCACCCC GCTGCTGGCC
 7741 GCCCTGGTGA TCTGCTACGT GGGCGGTGTC GTCTCGGCCT TCGCCTCGAC CACCGGGATC
 7801 CTCGGTCCCC TGATGCCGCT GTCCGAGCCG TTCCCTGAAGT CCGGTGCCAT CGGGACGACC
 7861 GGCATGGTGA TGGCCCTGGC GGCGCGGCG ACCGTGGTGG ACCCGAGTCC CTTCTCCACC
 7921 AATGGTGCTC TGGTGGTGGC CAACGCTCCC GAGCGGCTGC GGCGCGGCGT GTACCAAGGGG
 10 7981 TTGCTGTGGT GGGCGCCCG GGTGTGCGA CTGGCTCCC CGGCCGCCTC GGCGGCCTTC
 8041 GTGGTGGCGT GAGCGCAGCG GAGCGGGAA CCCCTGGAGC CCCTTCCCC TGCTGTGTCG
 8101 CTGACGTAGC GTCAAGTCCA CGTGGCGGGC GGGCAGTAGC CCTAGCATGT CGGGCATGGC
 8161 TAATCAGATA ACCCTGTCCG ACACGCTGCT CGCTTACGTA CGGAAGGTGT CCCTGCGCGA
 8221 TGACGAGGTG CTGAGCGGC TGCGCGCGA GACGGCCGAG CTGGCGGGCG GTGGCGTACT
 15 8281 GCCGGTGCAG GCGGAGGAGG GACAGTCTCT CGAGTTCCTG GTGCGGTGCA CGGGCGCGCG
 8341 TCAGGTGCTG GAGATCGGG ACGTACACCGG CTACAGCAGC CTCTGCCCTGG CCCGCGGATT
 8401 GGCGCCCGGG GGCGTGTGG TGACGTGCGA TGTATGCCG AAGTGGCCCG AGGTGGCGA
 8461 GCGGTACTGG GAGGAGGCCG GGGTTGCCGA CGGATCGAC GTCCGGATCG GCGACCCCCG
 20 8521 GACCGTCCTC ACCGGGCTGC TCGACGAGGC GGGCGCGGGG CGCGAGTCGT TCGACATGGT
 8581 GTTCATCGAC GCGACAAGG CGGGTACCCC CGCTACTAC GAGGGCGCGC TGCGCTGGT
 8641 ACGCCGCGGC GGGCTGATCG TCGTCGACAA CACGCTGTT TTCCGGCCGGG TGGCGACGA
 8701 AGCGGTGCAG GACCCGGACA CGGTCGCGGT ACAGCGAACTC AACCGGGCAC TGCGCGACGA
 8761 CGACCGGGTG GACCTGGCGA TGCTGACGAC GGCGGACGGC GTCACCCCTGC TGCGGAAACG
 8821 GTGACCGGGG CGATGTCGGC GGCGTGCAGC GTCAAGTCG TGCGCGCGGG CCTCGCGGAG
 25 8881 GGCTCCAGAT GCAGGGTTC GACGCCGGCG GCGGAAGCGC CGGCCACCTC GGACACCGAG
 8941 GGGCAGTCGG AGTCCGCGAA GCCCCCGAAC CGGTAGGCGA TCTCCATCAT GCGGTTGCGG
 9001 TCCGTACGCC GGAAGTCCCG CACCAAGGTGC GCCCCCGCGC GGGCGCCCTG GTCCGTGAGC
 9061 CAGTTCAAGGA TCGTCGCACC GGCACCGAAC GACACGACCC GGCAGGACGT GGCGAGCAGT
 9121 TTCAGGTGCC ACAGTCGACGG CTTCTCTCC AGCAGGATGA TGCCGACGGC GCCGTGGGG
 30 9181 CCGAAGCGGT CGCCCATGGT GACGACGAGG ACCTCATGGG CGGGATCGGT GAGCACCGC
 9241 GCAGGTGGC GTCGGAGTAG TGACGCGCG TCGCGTTCAT CTGGCTGGTC CGCAGCGTC
 9301 GTTCCTCGAC GCGGCTGAGT TCCTCCTCCC CGCGGGTGC GATCGTCATG GAGAGGTGCA
 9361 GCGAGCGCAG GAAGTCTCG TCGGGACCGG AGTACGCCCT CGGGGCTGG TCGCGCGCGA
 9421 AACCCGCCCTG GTACATCAGG CGCGGCCGAC GCGAGTCGAC CGTGGACACC GGCGGGCTGA
 35 9481 ACTCCGGCAG CGACAGGAGC GTGGCCGCCT GCTCGGCCGG GTAGCACCGC ACCTCGGGCA
 9541 GGTGGAACGC CACCTCGGCA CGCTCGGCCGG GCTGGTCGTC GATGAACCGC ATCGTGGTCG
 9601 GTGCAGATT CAGCTCCGTG GCGATCTCGC GGACGGACTG CGACTTCGGC CCCCATCCGA
 9661 TCGGGGCCAG CACGAAGTAC TCCGCCACAC CGAGGCCTTC CAGACGCTCC CACCGAGGT
 9721 CGTGGTCGTT CTTGCTCGCC ACCGCCTGGA GGATGCCGCG GTCGTCGAGC GTGGTGATCA
 40 9781 CCTCGCGGAT CTCGTCGGTG AGGACCACCT CGTCGTCTC CAGCACGGTG CCCCAGGACA
 9841 AGGTGTTGTC CAGGTCCCAG ACCAGACACT TGACAATGGT CATGGCTGTC CTCTCAAGCC
 9901 GGGAGCGCCA GCGCGTGTG GGCCAGCATC ACCCGGCACA TCTCGCTGCT GCCCTCGATG
 9961 ATCTCCATGA GCTTGGCGTC GCGGTACGCC CGTTGACGA CGTGTCCCTC TCTCGCCCT
 10021 GCGGACCGA GCAACCTGTGC GGCGTGCAGC GCCCCGGCGG CGGCTCGTT GCGGGCGACG
 45 10081 TGCTTGGCCA GGATCGTCGC GGGCACCATC TCGGGCGAGC CCTCGTCCCA GTGGTCGCTG
 10141 GCGTACTCGC ACACGCGGGC CGCGATCTGC TCCGCGTCC ACAGGTCGGC GATGTGCCCG
 10201 GCGACGAGTT GGTGGTCGCC GAGCGGCCGG CGAAGTGTCT CCCGGGTCCG GCGTGGGCC
 10261 ACCGCGCGGG TCGGGCAGGC CGCAGGATC CGACGCGACG CCCAGGCGAC CGACTTCGCG
 10321 CCGTAGGCCA GTGACGCCGC GACCAAGCATC GGCAGTGAAG CGCCGGAGCC GGCCAGGACC
 50 10381 GCGCGGCCG GCACACGAC CTGGTCCAGG TGCAGATCGG CGTGGCGGGC GCGCGGCAG
 10441 CGGAGCGCT TCGGGACGCC CTCGACGCCGT ACGCCGGGGG TGTCGGCGGG CACGACCA
 10501 ACCGACCCGG AACCATCCCTC CTGGAGACCG AAGACGACCA GGTGGTCCGC GTAGGCGCG
 10561 GCAGTCGTCC AGACCTGTG GCCGTCGACG ACAGCGGTGT CCCCCTCGAG CGAACCCGC
 10621 GTCCGCATCG CCGACAGATC GCTGCCCGCC TGCCGCTCAC TGAAGCCGAC GGCCCGAGT
 55 10681 TTCCCGCTGG TCAAGCTCCT CAGGAAGGTC GCGCGCTGAC CGGCGTCGCC GAGCGCTGC
 10741 ACGGTCCACG CGGCCATGCC CTGCGACGTC ATGACACTGC GCAGCGAACT GCAGAGGCTG
 10801 CCGACGTGTG CGGTGAACCT CGCGTCTCC CGGCTGCCGA GTCCCGAGACC GCGTGTCTG
 10861 GCGGCCACTT CCGCGCAGAG CAGGCCGTG GCGCGAGCC GGACGAGCAG GTCGCGCGGC
 10921 AGTCGCGGG ACAGTGTCCCA CTCGGCGGGC CGGTACCGA CAAGGTCGGT CAGCAGCGCG
 60 10981 TCACGCTCAG GCATCGACGG CCCGCGACGG GTGGACGAGT GCGACCATGG ACTCGACGGT

11041 ACGGAAGTTC GCGAGCTGGA GGTCCGGGCC GGCGATCGTG ACGTCGAACG TCTTCTCCAG
 11101 GTACACGACC AGTTCCATCG CGAACAGCGA CGTGAGGCCG CCCTCCGCGA ACAGGTCGCG
 11161 GTCCACGGGC CAGTCCGACC TGGTCTTCGT CTTGAGGAAC GCGACCAACG CGTGCACGAC
 11221 GGGTCTGTCC TTGACGGGTG CGGTCTATGAG AACACCTTCT CGTATTCTGTA GAAGCCCCGG
 5 11281 CCGGTCTTCC GGCGTGTGGT TCCCTCGCGG ACCTTGCCTA GCAGCAGGTC ACAGGGCGG
 11341 CT3CGCTCGT CGCCGGTGC G TTTGTCCAGC ACCCACAGCG CGTCGACGAG GTTGTGATG
 11401 CCGATCAGGT CGCGGGTGC CAGCGGCCCG GTCGGATGGC CGAGGCACCC CGTCATGAGC
 11461 GCGTCCACGT CCTCGACCGA CGCGGTGCCG TCCGTGACGA TCCCGCGCCG GTCGTTGATC
 10 11521 ATCAGGGTGA GCAGCCGGCT CGTGACGAAG CGGGGCGCGT CCCGGACGAC GATCGGCTTG
 11581 CGCCGCAGCG CGCGGAGCAG GTCCCCGGCG GCGGCATGG CCTTCTCACCG GGTCCGGGGT
 11641 CCGGGATCA CCTCGACCGT CGGGATCAGG TACGACGGGT TCATGAAGTG CGTGCAGGAC
 11701 AGGTCTCGG GCCGGGCCAC GGAGTGGGCC AGTCTGCTCAA CGGGGATCGA CGACGTGTT
 11761 GTGATGACCG GGATACCGGG CGCCGATGCC GAGACCGTGG CGAGTACCTC CGCCTTGACC
 15 11821 TCGGCGTCT CGACGACGGC CTCGATCACC GCGGTGGCG TACCGATCGC GGGCAGCGCG
 11881 GACGTGGCCG TCCGCAGCAC ACCGGGGTCG GCCTCGCGG GCGCCGCCAC GAGTTGTGCC
 11941 GTCCGCAGTT CGGTGGCGAT CGCGCCCGC GCCGCCGTA GGATCTCTC GGACGTGTCG
 12001 ACGAGTGTCA CGGGGACGCC GTGGCGCAGC GCGAGCGTGG TGATGCCGGT GCCCACACT
 12061 CCCGCAGCGA GCACGATCAG CTGGTGGTCC ACGCTGTTT CTCCTCCGG GGTACCATATG
 20 12121 GCAGCGAGTA CGGGTGCAGG ACGTCTTCCG GGGTCGACCC GATCGCTGCC TTGCGGCCGA
 12181 GGCCGAGTTC GTCGCGAAG CGGAGCAGCA CGTCGAAACG GATGTGGTGC CGGAACGCCG
 12241 TGCCCGTGA GTCGAGGACG CTCAGGCTGT CCCGGTGGTC CGCCGCGGTG TCCGGTGCCG
 12301 CGCACAGGGC CGCCAGGCC GGGCCGAGCT CGCGGTCGGG CAGTTGCTGG TACTCGCCCT
 12361 CGGCGCGGGC CTGCCCCGGG TGTCGACGC AGATGAACGC GTCGTCGAGC AGGGTCTTCG
 12421 GCAGTTCGGT CTTGCCCCGGC TCGTCGGCGC CGATGGCGTT CACATGCAGG TCGGGCAGCC
 25 12481 GCGGCTCGGC GGGCAGCACC GGGCCCTTGC CCGAGGGCAC CGAGGTGACG GTGGACAGGA
 12541 CATCCCGGGC GGCGCGGGCC TCCGCCGGAT CGGTACACTT GACCGGGCAGT CCGAGGAACG
 12601 CGATGCGGTC CGCGAACGAC GCCCGTGGC CGGGGTCGGT GTCGCTGACC AGGATCCGCT
 12661 CGATGGGCAG GACCTGCTG AGCGCGTGC CCGGGTCAC CGCCTGTGC CCGCGCCGA
 12721 TCAGCGTGAG CGTGGCGCTG TCGGACCGGG CCAGCAGCCG GTCGCGACG GCGGCACCG
 30 12781 CGCCGGTCCG CATCGCGGTG ATCACGCCG CGTCGGCAG GGGGGTCAGA CTGCCGCTGT
 12841 CGTCGTCGAG GCGCAGACATC GTGCCGACGA TCGTCGGCAG CCGGAAGCGC GGATAGTTGT
 12901 GCGGACTGTA CGAAACCGTC TTCATGGTCA CGCCGACACC GGGGACCCGG TACGGCATGA
 12961 ACTCGATGAC GCCGGGAATG TCGCCGCCGC GGACGAATCC GGTACGCGGC GGCACCTCGG
 13021 CGAACTCGCC GCGGCCGAGC GCGCGAACCG CGTCGTGAC TCGCTGATC AGCCGGTCCA
 35 13081 TCATCACGTC GCGGCCGATC ACGGAGAGAA TCCGCTTGAT GTCACGTTGG CGCAGGACCC
 13141 TGGTCTGCAT GTGTCACCTC CCTTCTGTGG CGGGAGCTGT CTGGTGGTGC CCGCTCGGGG
 13201 CGGCTCCGT TCTCATCGCA GCTCCCTGTC GATGAGGTG AAAATCTCGT CGCGGGTCGC
 13261 GTCCCGGGAC AGCACGCCG CGGGCGTGGT CGGGCGGGT TCCCAGCCGCC AGCGGTTGAG
 13321 CAGGGCGTCC AGCCGGGTC CGATCGCGC CGCCTGGCG GCGCCCGGGT CGACACCGGC
 40 13381 AACGAGTGCT TCCAGCCGGT CGAGCTGCGC GAGCACACG GTCACCGGGT CGTCCGGGA
 13441 CAGCAGTTCA CCGATCGGGT CGGGCGAGTGC GCGCGCGAC GGGTAGTCGA AGACGAGCGT
 13501 GGCAGACAGT CGCAGACCGG TCGCCTCGTT GAGGCCGTT CGCAGCTGCA CGCGCATGAG
 13561 CGAGTCCACA CCGAGTTCCC GGAACGCCGC TCCCTCCGGG ATGTCCTCCG GTTCGGCGTG
 13621 GCCCAGGACG GCGCTGCGCT TCTGCCGGAC GAGGGCGAGC AGGTGGTGG GCGTCTCGT
 45 13681 CTCGTGGCG GCGCTCCGGC GGGCCGACGG CTTGGGCCGG CACCGCAGCA GCGGGAGGTC
 13741 CGGCGGCAGG TCGCCCGCA CGGGCACGAC ACTGCCGTT CCGGTGTGGA CGCGGGCGTC
 13801 GTACATGCGC ATGCCCTGTT CGGGCGTGGAG CGCGCTCGCC CAACCCCTTCG GCATACGGCG
 13861 CGGGTCCGGC TCGGTCAAGGT CGGGCGTCAG GCAACTCGCC TGGTCCCACA GCCCCCACGC
 13921 GATCGACAGC CCTGGCAGCC CTTGTCGACCG CCGGTGTTCG GCGAGGCCGT CGAGGAACGC
 50 13981 GTTCGCCGCC CGCTAGTTGC CCTGACCGGG GGTGCCCCAGC ACACCGGCCG CGCAGCAGTA
 14041 GACGACGAAT CGGGCGAGGT CGGTGTCGCG GGTGAGCCGG TGCAAGGTGCG AGGCAGCGTC
 14101 GGCCTGGGT TTGAGGACGG TGTCGATGCG GTCGGGGGTG AGTTGTCGA GCAGGGCGTC
 14161 GTCGAGGGTT CGGGCGGTG GGAAGACGGC GGTGAGGGGT TGAGGGATGT GGGCGAGGGT
 14221 GGTGGCGAGT TGGTGGGGT CGCGCACGTC GCAAGGGAGG TGGGTGCCGG GGGTGGTGTG
 55 14281 GGGGGGTGGG GTGCGGGAGA GGAGGTAGGT GTGGGGGTGG TTCAGGTGGC GGGCGAGGAT
 14341 GCGGGCGAGG GTGCCGGAGC CGCGGGTGT GACGACGGCC CCGTCGGGGT CCAGCGGCCG
 14401 CGGGACCGTG AGGACGATCT TGCCGCTGTC CTGCGCGG CTCATGGTCG CCAGCGCCTC
 14461 GCGGACCTGC CGCATGTGTC GCACCGTCAC CGGCAGCGGG TGCAAGCACAC CGCGCGCGAA
 14521 CAGGCCGAGC AGCTCCGCGA TGATCTCCTT GAGCCGGTGC GGGCCCGCGT CCATCAGGTC
 60 14581 GAACGGTCGC TGGACGGCGT GCCGGATGTC CGTCTCCCC ATCTCGATGA ACCGGCCACC

14641 CGGCGCGAGC AGGCCGACGG ACCCGTCGAG GAGTTCACCG GTGAGCGAGT TGAGCACGAC
 14701 GTCGACCGGC GGGAACGCGT CGGCGAACGC GGTGCTGCGG GAATCGGCCA GATGCGCTCC
 14761 GTCCAGGTCC ACCAGATGGC GCTTCGCGC GCTGGTGGTC GCGTACACCT CCGCGCCCAG
 14821 GTGCCCGCGC ATCTGCCGGG CGGCGGAACC GACACCGCCG GTGGCGCGT GGATCAGGAC
 5 14881 CTTCTCGCCG GGGCGCAGGC CGGCGAGGTC GACCAGGCCG TACCAACGCGG TCGCGAACGC
 14941 GGTCATCACG GACGCCGCT GCGGGAACGT CCAGCCGTC GGATCCCGC CGAGCATCCG
 15001 GTGGTCGGCG ATGACCGTGG GGCGGAAGCC GGTGCGACG AGGCCGAAGA CGCGGTCGCC
 15061 CGGTGCCAGA CGGGAGACGT CGGCGCCGGT CTCCAGGACG ATGCCCGCGG CCTCGGCCGCC
 15121 GAGCACGCCG TGACCCGGGAG AGGTGCCGAG CGCGATCACG ACATCGCGGA AGTTGAGGCC
 10 15181 CGCCGCACGC ACACCGATCC GGACCTCGGC CGGGCGAGG GGGCGCCGGG GCTCCGCCGA
 15241 GTGGCCCGCG GTGAGGCCGT CGAGGGTGCG CGTCCGCGCC GGCGGATCA GCCACGTGTC
 15301 GCTGTCCGCC ACGGTGAGCG GCTCCGGCAC CGGGGTGAGG CGGGCCGCCCT CGAACCGGCC
 15361 CGCGCGCAGC CGCAGACGCCG GCTCCCGAG TGGCACGGCG ATGCCGTGCT GCTCGGGGGC
 15421 GAGCGTGACG CGCGACTCGG TCTCGACGTG GACGAACGG CGGGGCTGCT CGGCCTGGGC
 15 15481 GGCGCGCAGC AGTCCGGCCG CGCGCCCGGT GGCGAGGGCC CGGGTGGTGT GCACGAGCAG
 15541 ATCCCCGCCG GAGCCGGTCA GGGCGGTCA CGAGCCGGT CAGCCGGGTG GTGAGCGCAC
 15601 CACCGGGTCA TCGCCATCAG CGGCAGGCAA CGTGTGACG TCCACGTGCG GCGTCTGGC
 15661 ATCCGTGGGT GCGGCACCT CGATCCAGGT GAGACGCATC AGGCCGGTGC CGACGGGTGG
 20 15721 GGACAGCGGG CGGGTGCAGA CGTCCGGAT CTCGGCGACG AGTGGATCACG GCTCGGAGCA TGGCCGAGCC
 15781 GACGCGCAGA CTCAGCTCGT CGCCGTCACG AGTGGATCACG GCTCGGAGCA TGGCCGAGCC
 15841 CGTGGCGACG AACCGGGCCC CCTTCCAGGC GAACGGCAGA CCCCGAGCGC TGTGGTCCGG
 15901 CGTGGTGAGG GCGACGGCGT GCAGGGCCGC GTCGAGCAGC GCGGATGCA CACCGAAACC
 15961 GTCCGCTCG GCGGCCTGCT CGTGGCGAG CGCCACCTCG GCATACACGG TGTACCATC
 16021 ACGCCAGGCA GCCCCGCAACC CCTGGAACGC CGACCCGTAC TCATAACCGG CATCCCGAG
 25 16081 TTCGTCTAG AACCCCCGAGA CGTCGACGGC CACGGCCGTG ACCGGCGGCC ACTGCGAGAA
 16141 CGGCTCCACA CGCACAAACAC CGGGGGTGTG GGGGGTGTG GGGGGTCAAGGG TGCCGCTGGC
 16201 GTGCCGGGTG CAGCTGCCCG TGCCCTCGGT ACAGCGCTGG ACGGTCACCG GCGCCGGTCC
 16261 GGCCTCATCA GCCCCTTCCA CGGTACCGA CACATCCACC GCTGCGGTCA CGGGCACCAAC
 16321 AAGGGGGGAT TCGATGACCA GTCGTCAC TATCCCGAA CGGGTCTCGT CACCGGGCCG
 30 16381 GATGACCGAGC TCCACAAACG CGTACCCGG CAGCAGGACC GTGCCCGCA CGCGTGATC
 16441 AGCCAGCCAG GGGTGAGTGC GCAATGAGAT CGGGCAGTG AGAACAAACAC CACCATCGTC
 16501 GGCGGGCAGC GCTGTGACAG CGGCCAGCAT CGGATGCGCC GCACCCGTCA ACCCCGCCGC
 16561 CGACAGATCG GTGGCACCGG CGCCTCCAG CGAGTACCGC CTGTGCTCGA ACGCGTACGT
 16621 GGGCAGATCC AGCAGCCGTC CGGGCACCGG TTGACCACCG GTGCCCCAGT CCACTGCCGT
 35 16681 GCCCCAGGGTC CACGCGCTCG CCAACGCCGT CAGCCACCGC TCCAGGCCG CGTCACCGGT
 16741 CCGCAACGAC GCCACCGTGT GAGCCTGCTC CATCGCCGGC AGCAGCACCG GATGGGACT
 16801 GCACTCCACG AACACCGACCATCCAGCTC CGCCACCGCC GCGTCCAACG CCACCGGACG
 16861 ACGCAGATTG CGGTACCAAGT ACCCCCTCATC CACCGGCTCC GTCAACCCAGG CGCTGTCCAC
 16921 GTTCGACCAAC CACGCCACCG ACGCGCCCTT CCCTGCCACC CCCTCCAGTA CTTGGCCAG
 40 16981 TTCATCCTCG ATGGCTTCCA CGTGGGGCGT GTGGGAGGGG TAGTCGACCG CGATACCGACG
 17041 CACCCGACG CCTTCGGCCCT CATAACCGC CACCAACCTCC TCCACCGCCG ACGGGTCCCC
 17101 CGCCACCAAC GTCGAAGCGG GGCGTTACG CGCCGCGATC CACACACCC CGACCAGACC
 17161 GACCTCACCG GCCGGCAACCG CCACCGAAGC CATCGCTCCC CGCCCGGCCA GTCGCGCCGC
 17221 GATGACCTGA CTGCGCAATG CCACCAACGCG GGCAGCGTCC TCGAGGCTGA GGGCTCCGGC
 45 17281 CACGCACGCC GCCCGCATCT CGCCCTGGGA GTGTCCGATC ACCGGCGTCG GCACGACCCCC
 17341 ATGCGCTGTC CACAGCGCGG CGAGGTCAC CGCGACCGCC CAGCTGGCG GCTGGACCAC
 17401 CTCCACCCGC TCCGCCACAT CGGCGCGC CAACATCTCC CGCACATCCC AGCCCGTGTG
 17461 CGGCAGCAAC GCGTGGCGC ACTCCCTCCAT ACGCGCGGG AACACCGCGG AGTGGGCCAT
 17521 GAGTCCACG CCCATGCCGA CCCACTGGGC GCGCCGGCG GGGAAAGACGA ACACCGTACG
 50 17581 CGGCTGGTCC ACCGCCACAC CGTCACCCGG GGCATGCCCG AGCAGCACCG CACGGTGACC
 17641 GAAGACAGCA CGCTCCCGCA CCAACCCCTG CGCGACCGCG GCCACATCCA CACCAACCCCC
 17701 GCGCAGATAC CCCCTCCAGCC GCTCCACCTG CCGCCGGAGA CTCACCTCAC CACGAGCCGA
 17761 CACCGGCAAC GGCACCAACCG CGTCAACAAAC CGACTCCCCA CGCGACGGCC CAGGAACACC
 17821 CTCAAGGATC ACCTGCGCGT TCGTACCGCT CACCCCGAAC GACGACACAC CCGCATGCCG
 55 17881 TGCCCGATCC GACTCGGGCC ACGGCCCTCGC CTGGTGAGC AGCTCCACCG CACCGGCCGA
 17941 CCAGTCCACA TGCAGCGACG GCTCGTCCAC ATGCGCGTC TTGCGCGCGA TCCCGTACCG
 18001 CATGCCATG ACCATCTTGA TCACACCGCC GACACCCGCC GCGCCCTGCCG CATGACCGAT
 18061 GTTCGACTTC AACGAACCCA GCAGCAGCGG AACCTCACCG TCCCTGCCGT ACGTGCGCAG
 18121 AATGGCCCTGC GCCTCGATGG GATCGCCCGAG CGTCGTCCCC GTCCCGTGC CGTCCACAC
 60 18181 GTCCACATCG CGGGCGCGCA GTCCGGCGTT CACCAACGCC TGCTGGATGA CACGCTGCTG

	18241	GGACGGGCCG	TTGGGGCCGG	ACAGCCCGTT	GGAGGCACCG	TCCCTGGTTCA	CCGCCGACCC
	18301	GCGGACGACC	GCGAGAACCG	TGTGTCCGTT	GCGCTCGCG	TGCGAGAGCC	GCTCCAGCAC
5	18361	AAGAACGCCG	GCGCCCTCCG	CCCAGCCGGT	GCCGTTGGCG	GCGTCCGCGA	ACGCGCGGCA
	18421	GCGGCCGTG	GGGGAGAGTC	CGCCCTGCTG	CTGGAATTCC	ACGAACCCGG	TCGGGGTCGC
10	18481	CATGACGGTG	ACACCGCCGA	CCAGCGCCAG	CGAGCACTCC	CCGTGGCGCA	GTGCGTGC
	18541	GGCCTGGTGC	AGCGCGACCA	GCGACGACGA	GCACGCCGTG	TCCACCGTGA	ACGCCGGTCC
	18601	CTGGAGCCC	TAGAAAGTACG	AGATCCGGCC	GGTGAGCAGC	CTGGGCTGCA	TGCCGATCGA
	18661	GCCGAACCCG	TCCAGGTCCG	CGCCGACGCC	GTACCCGTAC	GAGAAGGCAGC	CCATGAACAC
15	18721	GCCGGTGTG	CTGCCCGCGA	GTGTGCCCGG	CACGATGCC	GCGCTCTCGA	ACGCCTCCCA
	18781	TGTCGTTTCC	AGCAGGATCC	GCTGCTGGGG	GTCCATGCC	CGTGCCTCAC	GGGGGCTGAT
	18841	GCCGAAGAAC	GCGGCATCGA	AGCCGGCGGC	GTCGGAGAGG	AAGCCGCCGC	GGTCCGTGTC
	18901	CGATCCGCCG	GTGAGGCCGG	ACGGGTCCA	GCCACGGTCG	GCCGGGAAGC	CGGTGACCGC
	18961	GTCGCCGCCA	CTGTCCACCA	TGCGCCACAG	GTCGTGGGC	GAGGTGACGC	CGCCCGGCAG
20	19021	TCGGCAGGCC	ATGCCACCGA	TGGCCAGCGG	TTCGTACCG	GTCCGGCG	CTGTGGAAC
	19081	AGCGACCGG	GCGGCACCAC	CGACCAAGAGC	CTCGTCCAAC	CGCGACGCCA	TGGCCCGCGG
	19141	CGTCGGTAG	TCGAAGACAA	GCCTGGCGGG	CAGTCGGACA	CCGTCGCG	CGGCAGTCG
	19201	GTTCCCGAGT	TGACGGCGG	TCAGCGAGTC	GATAACCCAGT	TCCTTGAAGG	CCCGCTCCGC
	19261	GGACACGTCC	GCGGCCTCCG	CGTGGCCGAG	CACCGCCGCC	GCGTTGTCG	GGACCAAGTC
25	19321	CAGCAGCGCG	GTGTCCCCTG	CAGCGCCGG	CATGGTGGCG	AGCCGGTCGG	CGAGCGAAC
	19381	GGCGGTGGCC	GCCGCCGGGG	GCGATACGGC	GCAGCGCAGA	TCGGCGAAAA	CGGGCGATGT
	19441	GTGCGCGGTG	AGGTCCATCG	TGGCCGCCAC	GGCGAACCGC	GTGCCGGTTC	CGGCCGCGGC
	19501	TTCCAGCAGG	CGCATGCCA	CACCGCCGA	CATGGGGCGG	AAACCGCCGC	GGCGGACACG
	19561	GGTGCGGTTG	GTGCCGCTCA	TGCTGCCGGT	GAGTCGGCTG	TCATCGGCC	AGAGGCCCA
30	19621	GGCCAGCGAC	AGCGCGGGCA	GTCCTCGGC	ATGGCGCAGC	GTCCGAGTC	CGTCGAGGAA
	19681	CCCGTCGCC	GCCGAGTAGT	TGCCCTGGCC	GGGGCGGCC	ATGATGCCG	CGACGGACGA
	19741	GTAGAGGACG	AACGAGCGCA	GGTCCCGCTC	CCGGGTCAGC	TCGTGCAGGT	GCCAGGCAGC
	19801	GTCGGCTTTG	GGGCGCAGTG	TGGTGGCGAG	CCGCTCCGGG	GTGAGTGC	TGGTCACGCC
	19861	GTCGTGAGC	ACGGCTGCCG	TGTGGAAGAC	CGCCGTGAGC	GGCCTGCCGG	CGGCGGGAG
35	19921	CGCGGGCGG	AGCTGGTCCC	GGTCGGCGAC	GTCACAGCGG	ATGTGGACAC	CGGGAGTGT
	19981	CGCCGGCGGT	TCGCTGCCG	ACAGCAACAG	GAGGTGGCGG	GGCGCATGCT	CGGCAGCGAG
	20041	ATGCCGGCG	AGGAGACCTG	CCAGCACACC	CGAGCGCCG	GTGATGACCA	CCGTGCCGTC
	20101	CGGGTCGAGC	ACGGGTTCCG	GCGTTCCGC	GGCGGCCGTG	CGGGTGAACC	CGGGCGCTTC
	20161	GTACCGGCCG	TCGGTGACCC	GGACGTACCG	CTCGGCCAGT	GTCCGGCGG	CGGCCAGCGC
40	20221	CTCGATGGGG	GTGTGGTGC	CGGTCTCCAC	CAGCACGAA	CGGCCCCGGT	GCTCGGCTG
	20281	GGCGGACCGG	ACGAGGCCGG	CGACCGCTCC	TCCGACCGG	CCCGCGTCGA	TCCGGACGAC
	20341	GAGGGTGGTC	TCCGCAGGGC	CGTCCTCGGC	GATCACCCGG	TGCACTCGC	CGAGCACGAA
	20401	CTCGGTGAGC	CGGTACGTCT	CGTCGAGGAC	ATCCGCGCCC	GGTTCGGGA	CGCGGGAGAC
	20461	GATGTGGACC	CGGTCCCGAG	GACCGGGCCC	GGGAGTGGC	AGCTCGGTCC	AGGAGAGGCC
45	20521	GTACAAGGAG	TTCCGTACGA	CGGCGCGTC	GGCGTCGACG	TTCACCGGTC	CGCGGGTCAG
	20581	CGCGGCCACG	GTCACCAACCG	GTTGGCGAC	CGGGTCCGTC	GCATGCACGG	CAGCGCCGTC
	20641	CGGGCCCTGA	GTGATCGTGA	CGCGCAGCGT	GGTGGCCCCG	GTCTGTGGA	ACCGCACGCC
	20701	GCTCCACGAG	AACGGCAGCC	GCACCTCCGC	TTCCCTGTTCC	GCGAGCAGCG	GCAGGCAGGT
	20761	GACGTGCAAG	GCCGCGTCGA	ACAGCGCCGG	GTGGACGCCA	TAGTGC	TGTCGTCGC
50	20821	CTGTTCCCCG	GCGATCTCCA	CCTCGCGTA	CAGGGTTTCG	CCGTCGCGCC	AGGCGGTGCG
	20881	CAGTCCCTGG	AACGCTGGGG	CGTAGCTGTA	GCCGGTCTCG	GCCAGCGCT	CGTAGAACGC
	20941	GCTCACCGTC	ACCGCTCGCG	CGCCCCGGCG	GGGCCACCGC	GGCGCGGGGA	CCGCCGCGAC
	21001	GCTTCCGGCC	CGGCCGAGGG	TGCCGCTGGC	GTGCCGGGTC	CAGCTGTCCG	TGCCCTCGGT
	21061	ACGCGCGTGG	ACGGTCACTC	GCCGCCGTCC	GCCCTCATCG	GCCCGTTCGA	CGGTACCGA
55	21121	CACATCCACC	GCGCCGGTCA	CGGGCACAC	GAGCGGGGTC	TCGATGACCA	GTTCATCCAC
	21181	CACCCCGCAA	CCGGTCTCGT	CACCGCCCG	GATGACCAAGC	TCCACAAACG	CCGTACCCGG
	21241	CAGCAGAAC	GTGCCCGCA	CCGCGTGTAC	AGCCAGCCAG	GGATGCGTAC	GCAACGAGAT
	21301	CCGGCCAGTG	AGAACAAACAC	CACCAACGTC	GTCGGCGGGC	AGTGTGTGA	CGGCGGCCAG
	21361	CATCGGATGC	GCCGCCCGG	TCAGCCCGC	CGCGGACAGA	TCGGTGGCAC	CGGCCGCCTC
	21421	CAGCCAGTAC	CGCCTGTGCT	CGAACCGCGTA	GTTGGCAGA	TCGAGCAGCC	GTCCCGGCAC
60	21481	CGGTCGACC	ACCGTGTCCC	AGTCCACTGC	CGTGCCAGG	GTCCACGCCT	GCGCCAACGC
	21541	CGTCAGCCAC	CGCTCCCGAC	CGCCGTCACC	GGTCCGCAAC	GACGCCACCG	TGTGAGCCTG
	21601	TTCCATCGCC	GGCAGCAGCA	CCGGATGGGC	GTCGACTCC	ACGAACACGG	ACCCGTCCAG
	21661	CTCCGCCACC	GCCGCGTCCA	GCGCGACGGG	GCGACGCAAG	TTCCGGTACC	AGTAGCCCTC
	21721	ATCCACCGGC	TCGGTCACCC	AGGCGCTGTC	CACCGTGGAC	CACCAAGGCCA	CCGACCCGGT
	21781	CCCGCCGGAA	ATCCCCCTCCA	GTACCTCGGC	CAACTCGTCC	TCGATGGCTT	CCACGTGGGG

21841 CGTGTGGGAG GCGTAGTCGA CCGCGATAcg GCGCACTCGC ACGCCTTCGG CCTCGTACCG
 21901 CGTCACCACT TCTTCCACCG CGGACGGGTC CCCCGCCACC ACAGTCGAAG ACGGGGCGTT
 21961 ACGCGCCGCG ATCCACACGC CCTCGACCGAG GTCCACCTCA CCGGCCGGCA ACGCCACCGA
 22021 AGCCATCGCC CCCGCCCGG CCAGCCGCC GGCGATCAC TGGCTGCGCA AGGCCACAC
 5 22081 CGGGCCGGCG TCCTCAAGGC TGAGGGCTCC GGCCACACAC GCCGCCGCGA TCTCGCCCTG
 22141 GGAGTGTCCG ACCACCGCGT CGGGCACGAC CCCATGCGCC TGCCACAGCG CGGCCAGGCT
 22201 CACCGCGACC GCCCAGCTGG CGGGCTGGAC CACCTCCACC CGCTCCGCCA CATCCGGCCG
 22261 CGCCAACATC TCCCACAT CCCAGCCCGT GTGCGGAAAC AACGCCCGC CACACTCCTC
 22321 CATACTGAGCC CGAACACACG CAGAACACGC CATCAACTCC ACACCCATGC CCACCCACTG
 10 22381 AGCACCTGC CGGGGAAAGA CGAACACCGT ACGCGGCTGA TCCACCGCCA CACCCATCAC
 22441 CGGGGCATCG CCCAACAAACA CGGCACGGTG ACCGAAGACA GCACGCTCAC GCACCAACCC
 22501 CTGCGCGACC GCGGCCACAT CCACACCACC CCCGCGCAGA TACCCCTCCA GCCGCTCCAC
 22561 CTGCCCCCGC AGACTCACC CACTCCGAGC CGAACCCGGC AACGGCACCA ACCCATCGAC
 22621 AGCCGACTCC CCACCGCGACG GCCCGGGAAAC ACCCTCAAGG ATCACGTGCG CGTTCGTACC
 15 22681 GCTCACCCCCG AAAGCGGAGA CACCGGGCCG GCGCGGACGT CCCGCGTCGG GCCACCGCCG
 22741 CGCCTCGGTG AGCAGTCTCA CGCGCCCTC GGTCCAGTCC ACATGCGACG ACGGCTCGTC
 22801 CACATGCAGC GTCTTCCGGC CGATGCCATA CGGCATCGCC ATGACCATCT TGATGACACC
 22861 GCGACACCC GCAGCCGCC CGCGATGACC GATGTTCGAC TTCAACGAAC CCAGCAGCAG
 22921 CGGAACCTCA CGCTCCTGCC CGTACGTCGC CAGAATCGCG TGCCGCTCGA TGGGATCGCC
 20 22981 CAGCGTCGTC CCCGTCCCGT GCGCCTCCAC CACGTCCAGC TCAGGGGGGG CGAGCCCCCGC
 23041 CTTGTGGAGG GCCTGGCGA TGACCGCTG CTGGGAGGGG CGGTTGGGTG CGGAGATGCC
 23101 GTTGGAGGC CGCTCCTGGT TGACGGCGGA GGAGCGGACG ACCCGAGGAGA CGGTGTGTCC
 23161 GTTGCCTCG GCGTCGGAGA GCTTTTCGAC GACGAGGACG CGGGCCCCCT CGGCAGAAC
 23221 GGTGCCGTCC GCGCGTCAG CGAACGCCCT GCACCGTCCG TCCGGCGCGA CGCCGCCCTG
 25 23281 CGGGGAGAAC TCCACGAAGG TCTGTGGTGA TGCCATCACT GTGACACCAC CGACCAGCGC
 23341 CAGCGAGCAC TCCCCGGTCC GCAGCGCTG CCCGGCTGG TGCAGCGCGA CGAGCGACGA
 23401 CGAACACGCC GTGTCGACCG TGACCGCCGG ACCCTCCATG CGGAAGAAAGT ACGACAGCCG
 23461 TCCGGCGAGC ACCGCGGGCT GTGTGCTGTA GGCGCCGAAT CGGGCCAGGT CGCGCCCGT
 23521 GCCGTAGCCG TAGTAGAACG CGCCGACGAA GACGCCGGT CGCTGCGCGC GCAGGTGTC
 30 23581 CGGCACGATG CGGGCGTGT CGAGCGCTC CCAGGCATT TCGAGGAGGA TCCGCTGCTG
 23641 CGGGTCGAGT CGGGTGGCCT CGCGCGGACT GATGCCGAAG AACCGGGCAT CGAACGTCGGC
 23701 GGCGCCCGCG AGTGCGCCGG CCCGCCCGGT GGCGGACTCG GCAGCGCCGGC GCAGCGCCGG
 23761 CACGTCCCAG CGCGGGTCCG TGGGGAAGTC CGCGATCGCG TCGGGGCCGT CGCGACGAG
 23821 CTGCCACAGC TCTTCCGGT AGGTGACGCC GCCCGGCACT CGGCAGGCCA TGCCGACGAC
 35 23881 GGCAGCGGC TCGTTGCCG CGGCGCGAG CGCGGTGTT TCCGGCGGA GCTGCGCGTT
 23941 GTCCTTGACC GACGTCCGCA GCGCCTCGAT CAGGTCGTT TCAGGCCATCG CCTCATCCCT
 24001 TCAGCACGTG CGCGATGAGC GCGTCTGCGT CCATGTCGTC GAACAGTTCG TCGTCCGGCT
 24061 CCGCGGTCTG GGTGCTCGCG GGTGCCTGTC CGCGTGGTTC ACCGCCGTCC GGGGTCCCGT
 24121 TGTGTCGGG GGTCCCGTTG ACGTCCGGGG CGAGGAGGGT CAGCAGATGA CGGGTGAGCG
 40 24181 CGCCGGCGGC GGGATAGTCG AAGACGAGCG TGGCCGGAG CGGAATGCCG AGGGCCTCGG
 24241 AGAGCCGGTT CGCGAGGCCG AGCGCGGTGA GCGAGTCGAC CCCGAGGTCC TTGAACGCCG
 24301 TGGTGGCGT GACCGCCGCC GCGTCGGTGT GGCCCAGCAG GGTGGCGCG GGTGCGCGGA
 24361 CGACGCCGAG CAGCACCTGT TCCCGTTCCCT TGTGGGGCAG GTCCGGCAGG CGTTCCAGCA
 24421 GGGAGCCGCC GTGGTCCGCG GAGGCCCGGG TGGGGCGCTG GATCGGTGCG CACAGCGGTG
 45 24481 ACGGGTCGCC GGGCCCGGGT GGGGCGGTG CGACGACAC GGCTTCCCGT GTGGCGCACG
 24541 CGGCGTCGAG GAGGTGCGTC AGCCGGTCCG CGCGCGCGGT GAACGCCACG GCGGGCAGGC
 24601 CTTGTGCCCG CGCGAGGTGCG GCCAGGGCCT GGAGCGGTCC GGCCGCTCG CGGGACGGAA
 24661 CGGCAGAAC GAACGCGGTC AGGTGAGGT CGCGGGTCAG GCGGTGCACT TCCCAGGCCG
 24721 ACTCGCCGGT GCCGTCCCGC TGGACGACCG CGGTACCCCG GGTTCCGGC ACTGTGCCCG
 50 24781 GTCGTACCG GATCACTTCG GCGCGTGTG CGCCGAGGTG TCCGGCGAGT TCCTCCGAAC
 24841 CGCCCGCGAG GAGGACGGTG TCGCCGTACG AGGCCGCCG CGTGGTGGGC GCGGCAGGGAA
 24901 CGAGGGGGGG CGCTTCGAGG CGCCCGTCGG CCAGGCCGAG GTGCGGTTCG TCGAGGCGGG
 24961 AGAGGGCGGC CGCGCGGCCGG GGGGTGACCG TGTCGGTGGT CTCCACGAGC ACGAGCCGGC
 25021 CGGGTCCGC GGTGTCGAGC AGTGCAGCGA CGGCACCGGC GACGGGCCCG GCCTCGCGG
 55 25081 ACACCACCAAG CGTGGCGGCC GCGGTCTCGG GTGCGTCCAG TGCGGTACGG ACCTCGTCGG
 25141 GACCGGATAC CGGGACGACG ATGACGTCGG CGTGGCGTC GTGCCCGAGG TCGGTGTACC
 25201 GGCGGGCCGT GGTGCGGGGT GCGGCCGGGG CCCGGACGCC GGTCCAGGTG CGCCGGAACA
 25261 CGCGCACGTC CCCGTCCGGG CCCGTGCGTGG CGGGGGGCCG GGTGATGAGC GAGCCGATCT
 25321 GAGCCACCGG CGGTCCCAAGT TCGTCCGCGA GGTGACGCG GGCGCCGCC TCGCCCTCGC
 60 25381 CGTGGACGAA GGTGACGCGC AGTTCTGTGG CGCCGCTGGT GTGGACACGG ACGCCGGTGA

	25441	ACCGAAGCGG	CAACCGTACC	CCCGCGTTCT	CGGCAGGCCGC	GCGATGCTG	CCCGCTTGCA
	25501	GCGCGGTGAC	GAGCAGGCC	GGGTGCAGTG	TGTAGCGGGC	GGCGTCCCTG	GCGAGGGCGC
	25561	CGTCGAGGGC	GACTTCGGCG	CAGACGGTGT	CTCCGTGGCT	CCACGCGGCG	GACATGCCGC
5	25621	GGAACTCGGG	GCCGAACTCG	TATCCCCGCT	CGTCGAGTCG	CTGGTAGAAAG	GCCGCGACGT
	25681	CGACCGGTTTC	CGCGTGCCTG	GGCGGCCAGG	GCCCCGGCGT	GGTGGCCGGT	TCGGTGGTGG
	25741	CGATGCCGGC	GAAGCCGGAG	GCGTGGCGGG	TCCATGTCCG	GTCGCCGTCC	GTCCGGGCGT
	25801	GGACCGCAC	GGCACGGCGT	CCGGTGTCTG	CGGGCGCGGC	GACGGTCACG	CGCACCTGGA
	25861	CGGCGCCGGT	GGCGGGCAGG	ACCAAGCGGTG	TCTCGACGAC	CAGTTCGTCG	AGCAGGTGCG
10	25921	AGCCCTGCCCTC	GTCGGCGCCG	CGTCCGGCCA	ATTCCAGGAA	GGCGGGTCCG	GGCAGCAGTA
	25981	CGGCCCGTC	GACGGAGTGA	CCGGCCAGCC	ATGGGTGGGT	GGCCAGCGAG	AACCGCCCGG
	26041	TGAGCAGCAC	CTCGTCGGAG	TCGGGGAGCG	CCACCGACGC	GGCGAGCAGC	GGGTGGTCGA
	26101	CGGCGTCGAG	TCCGAGGCCG	GAAGCGTCCG	TGCGGCCGCG	GGTCTCGATC	CAGTAGCGCT
	26161	CATGGTGGAA	GGCGTATGTG	GGCAGGTCGT	GTGCCGTGCG	CGTCGCGGGG	ACGACCGCCG
15	26221	CCCAGTCGAC	GGGCACGCCG	GTGGTGTGCG	CCTCGGCCAG	CGCGGTGAGC	AGCCGGTGGGA
	26281	CTCCCCCGCC	GGGGCGGAGC	GTGGCGACGG	TCGCGCCGTC	GATCGCGGGC	AGCAGCACGG
	26341	GGTGCACGCT	GACCTCGACG	AACACGGTGT	CACCCGGCTC	GGGGCAGCG	GTCACGGCCG
	26401	TGGCGAAGCC	TACGGGGTGG	CGCATGTTGC	GGAACCAAGTA	CTCGTCGTCG	AGCAGCGCGT
	26461	CGATCCAGCG	TTCGTCGGCG	GTGGAGAACC	ACGGGATCTC	GGCGTGTGCG	GAGGTGGTGT
20	26521	CCGCGACGAT	CCGCTGGAGT	TCGTCGTACA	GCGGGTCGAC	GAACGGGGTG	TGGGTGGGCG
	26581	AGTCGACGGC	GATGCGGCCG	ACCCAGACGC	CGCGGGCTC	GTAGTCGGCG	ATCAGCGTTT
	26641	CGACGGCGTC	CGGGCGCCCG	GCGACGGTCG	TGGTGGTGGC	GCCGTTGCGG	CCCGCGACCC
	26701	AGACGCCGTC	GATCCGGCG	GCATCCGCT	CGACGTCGGC	GGCCGGGAGC	GCGACCGAGC
	26761	CCATCGCGCC	GGCTCCGGCG	AGTTCGCGCA	GGAGCAGGAG	AACGCTGCGC	AGCGCGACGA
25	26821	GGCGGGCACC	GTCCTCCAGG	GTGAGCGCTC	CGGCGACACA	GGCCGCGGGCG	ATCTCGCCCT
	26881	GGGAGTGTCC	GATGACGGCG	TCCGGCGTA	CGCCCGGGC	CTCCCACACG	GCGGCCAGCG
	26941	ACACCATGAC	GGCCCAGCAG	ACGGGGTGC	CGACGTCGAC	GCGGCGGGTC	ACCTCCGGGT
	27001	CGTCGAGCAT	GGCGATGGGG	TCCCAGCCCG	TGTGCGGGAT	CAGCGCGTCG	GCGCATTGGC
	27061	GCATCCTGGC	GGCGAACACCC	GGGGAGGCCG	CCATCAGTTC	GACGCCCATG	CCGCGCCACT
30	27121	GCGGTCTTG	TCCGGGGAAAG	ACGAAGACGG	TGCGCGCTC	GGTAGCGGCC	GTGCCGGTGA
	27181	CGACGTCGTC	GTGAGCGACG	ACGGCGCGGT	GCGGGAACGT	CGTACGCGCTG	GCGAGCAGGC
	27241	CCGCGCGCAT	GGCGCGCCGG	TCGTGGCCGG	GACGGGGCGC	GAGGTGCTCG	CGGAGTCGGC
	27301	GGACCTGGCC	GTGAGGGCC	GTGGCGGTCC	GCGCCGAGAC	GGGCAGTGGT	GTGAGCGGGC
	27361	TGGCGATCAG	CGGCTCACCG	GGCTTCGAGG	CGGACGGCTC	CTCGGCCGCG	GGCTCCCCGG
35	27421	CCGGGTGGGC	TTCCAGCAGG	ACGTGGCGT	TGGTGCCTG	GACGCCGAAG	GAGGACACAC
	27481	CGCGCGCCCG	CGGGCGGTG	GTCTCGGGCC	AGGGCCGGGC	ATCGGTGAGG	AGTCGACCGG
	27541	CGCCGGCCGT	CCAGTCGACG	TGCGAGGACG	GCGTGTCCAC	GTGCAAGGGTG	CGCGCAGGG
	27601	TGCGGTGCCG	CATGGCGAGG	ACCATCTGA	TGACACCGGC	GACACCCGCG	GCGGCCTGAG
	27661	TGTGGCCGAT	GTTGGACTTC	AGCGAGCCCA	GCAGCACCGG	GGTGTGCGGC	CCCTGCCCGT
40	27721	AGGTGGCCAG	CACCGCCTGT	GCCTCGATGG	GATCGCCAG	CCTGGTGGCG	GTGCCGTGCG
	27781	CCTCCACGGC	GTCCACGTCC	GGCGGGGTGA	GCCCCGGCGT	GGCCAGGGCC	TGCCGGATCA
	27841	CCCCTCCTG	CGAGGGCCCG	TTCGCGCCG	ACAACCCGTT	GGAAAGCACCG	TCCTGGTTGA
	27901	CCGCCGAACC	CCGGACAAAC	GCCAGCACAC	GGTGGCCGTT	GCGCTCGGCA	TCGGAGAGCC
	27961	TCTCGACGAT	CAGCACACCG	GACCCCTCGG	CGAAAACCGGT	GCCGTCAAGCC	GCATCCGCGA
45	28021	ACGCCTTGCA	GGCGCGTCG	GGCGCGAGAC	CCCGCTGCTG	GGAGAACCTG	ACGAAGCCGG
	28081	ACGGCGAGGC	CATCACCGTG	ACGCCCGCA	CCAGGGCGAG	CGAGCATTG	CCGGAGCGCA
	28141	GTGACTGCC	GGCCTGGTGC	AGGCCACCA	GCGACGACGA	ACACGCCGTG	TCGACCGTGA
	28201	CCGCCGGACC	CTCCAGACCG	TAGAAGTACG	ACAGCCGACC	GGACAGCACA	CTGGTCTGGG
	28261	TGCCGGTCG	GGCGAAACCG	CCCAGGTCGG	TGCCGAGTCC	GTACAGCCCG	TCCACGTCCC
50	28321	CCATGAACAC	GGCGGTGTCG	CTTCCCGCGA	GCGACTCCGG	GAGGATCCCG	GGGTGTTCCA
	28381	GCGCTCCCCA	CGAGGTCTCC	AGGACCAAGAC	GCTGCTGCGG	GTCCATCGCC	AGCGCCTCAC
	28441	GCGGACTGTAT	CCCGAAGAAC	GCCCGCGTCG	AGTCCGCCAC	CCCGCGAGG	AAGCCACCAT
	28501	GACGCACGGT	CGACGTGCC	GGATGATCCG	GATCGGGATC	GTACAGCCCG	TCCACGTCCC
	28561	AACCACGGTC	CGTCGAAAC	GCCGTGATCC	CGTCACCAACC	CGACTCCAGC	AGCCGCCACA
55	28621	AGTCCTCCGG	CGACGCGACC	CCACCCGGCA	GCGGGCAGGC	CATCCCCACG	ATCGCCAACG
	28681	GCTCGCTCTG	CCGGACGGCC	GGGGTCGTGG	TGGGGTCGG	CGATGCCGTC	CGGCGGGACA
	28741	GCGCCCGGGT	GAGCTTCGCC	GCGACGGCGC	CGGGCGTCGG	GAAGTCGAAG	ACCGCGGTGG
	28801	CGGGCAGCCG	TACGCCCCGTC	GCCTCGGTGA	AGGCCTTGGC	CAGCCGGATC	GCCATGAGCG
	28861	AGTCGACGCC	GAGTCCCTG	AACGTGGCGG	TGCGCTCGAC	CCGTGCGGCC	CCGTCTGGC
60	28921	CGAGTACGGC	CGCGGTGCGAC	TGCGGGACGA	CGGCAGACAC	GTCCCTTTTCG	GCGTCCCGCG
	28981	CGGAGAGCCG	CGCGATCCGG	TCGGCGAGGG	TGGTGGCGCC	GGCCGCCCGG	CGCCGCCGGCT

29041 CCCGGCGCGG TCGCGCAGC AGGGCGAGC TGCCGAGGCC GGCGGGTCA GCGCGACCA
 29101 CGCGGGGTC CGAGGACCGC AACGCCGCT CGAACAGCGT CAGTCGCCT TCAGCGGTCA
 29161 GCGCGTCAC GCGTCGCGG CGCATGCGG CGCCGGTCA CGACGTCAGC CCGCTCTCG
 29221 GTTCCCACAG GCCCCAGGCC ACGGACAACG CGGGCAGTCC GGCTGCCCG CGCTCTCG
 5 29281 CCAGCGCGTC GAGGAACCGC TTGCGGGCG CGTAGTTGCC CTGTCGGGG CTGCCAGCA
 29341 CACCGCGCGC CGACGAGTAG AGGACGAACG CGGCCAGTTC CGTGTCTGG GTGAGTTCTG
 29401 GCAGGTGCCA CGCGGGTCA ACCTTCGGGC GCAGCACCGT CTCGAGCCGG TCAGGGGTGA
 29461 GCGCGGTGAG GACGCCGTC TCGAGGACGG CGCGGGTGT CACGACGGCC GTGAGCGGGT
 29521 GCGCGGGTCA GATCCCCGCC AGTACGGAGG CGAGTCTCTC CGCGTCGGCG ACGTCGCAGG
 10 29581 CGATCGCCGT GACCTCGGCC CGGGCACGT CGCTCGCCGT GCGCTGCGC GACAGCATCA
 29641 GCAGCGCGC CACGCCGTGG CGTTCGACGA GGTGGCGCT GATGATGCCG GCCAGCGTCC
 29701 CGGAGCCACC GGTGACGAGC ACGGTGCCGT CGGGTCTAG CGCCGGAGCG TCACCCGCCG
 29761 GGACGCCCGG GGCCAGACGG CGGGCTACA CCTGGCGCTC ACCGAGCACC ACCTGGGCT
 15 29821 CATCGAGCGC GGTGGCGCT GCGAGCAGCG GCTCGGGT GTCCGGGGCG CGTCGACGA
 29881 GGACGATCCG GCGGGGGTGT TCGGCTGCG CGGTCCGAC CAGTCCGGCG GCCCGGCCG
 29941 ACGCGAGACC GGGCCCGGTG TGGACGGCCA GGACCGCGTC GGCGTACCGG TCGTCGGTGA
 30001 GGAAGCGCTG CACGGCGGTG AGGACGCCGG CGCCCAGTTC CGGGGTCTCG TCGAGCGGGG
 30061 CACCGCCGCC GCCGTGCGCG GGGAGGATCA CCACGTCCGG GACCGTCGGG TCGTCGAGG
 20 30121 GGCGGTCTG CGCGGTCTG GCGGGCAGCT CGGGGAGCTC GGCCAGCACC GGGCGCAGG
 30181 GGCCCCGAAAC GGCTCCCGTG ATCGTCAGGG GGCGCTGCG CACCGCGCCG ATGGTGGCGA
 30241 CGGGCCCGCC GGTCTCGTC GCGAGGTGTA CGCCGTCAGC GGTGACGGCG ACGCTACCG
 30301 CGGTGGCGCC GGTGGCGTGG ACGGGACGT CGTCGAAACG GTACGGAAGG TGGTCCCCTT
 30361 CGCGGGCGAG GCGGAGTGG GCGCCGAGCA GCGCCGGGTG CAGGCCGTAC CGTCCGGCGT
 30421 CGGCGAGCTG TCCGTCGGCG AGGGCCACTT CCGCCCAGAC GGCGTCGTG TCGGCCAGA
 25 30481 CGGCGCGCGG GCGGGGCCAGC GCGGGCCCGT CGGTGTACCC GGCTCGGGCC AGACGGTCGG
 30541 CGATGTCGTC GGGGTCCACC GGCGGGCCCG TGGCGGGCGG CCACGTCGAC GGCATCTCCC
 30601 GCACGGCCGG GGGCGTCGCC GGGTCGGGGG CGAGGATTCC GTGCGCTGTC TCGGTCCACT
 30661 CCCCCGCCGC GTGCCGCGTG TGCACGGTGA CGCGCGGGCG GCGTCCGCC CGGGCGCGC
 30721 TCACCGTGAC GGAGAGCGCG AGCGCACCGG ACCGCGGCAG CGTGAGGGGG GTGTCCACGG
 30 30781 TGAACGTGTC GAGGGCGCCG CAGCGGCTT CGTCGCCCCC CGGGATCGCC AGATCCAGGA
 30841 GGGCCCGGGC GGGCAGCACC GCGAGGCCGT CAAGGGAGTG CGCCAGCGGA TCGGCGGCGT
 30901 CGACCCGGCC GGTGAGCACC AGGTGCCGG TGCCGGGAG GGTGACCGCC GCGGTCAAGCG
 30961 CGGGGTGCGC GACCGCGTC TGTCCGGCCG GGGCCGCCTC GCGCGGGTC TGGGTGCCGA
 31021 GCCAGTAGCG GACCCGCTCG AACGGGTACG TCGCGGGTG CGAGGCGCGT GCGGGCGCG
 35 31081 GTCGATGAC CTTCGGCCAG TCGACCGTGA CGCCGTCGGT GTGAGGCCGG GCGAGCGCGG
 31141 TCAGGGCGGA TCGCGGTTCG TCGTCGGCGT CAAGCATCGG GATGCCGTG ACGAGTCGGG
 31201 TCAGGCTCCG GTCCGGGCCG ATCTCCAGGA GAACCGCCCC GTCGTGCGC GCGACCTGTT
 31261 CCCCCAACCG GACGGTGTG CCGACCTGTC GTACCCAGTA CTCCGGCGTG GTGCAGGCCG
 40 31321 CGCCCGCGGC CATCGGGATC CTCGGCTCGT GGTACGTCAG GCTCTCCGCC ACCTTGCAGA
 31381 ACTCCTCGAG CATCGGCTCC ATCCGCGCCG AGTGGAACGC GTGGCTGGTC CGCAGGCCG
 31441 TGAAGCGGCC GAGCGGGCC GCGACGTCGA GAACCGCCTC CTCGTCACCG GAGAGCACGA
 31501 TCGACCGGGG CCCGTTGACC GCGCGATCT CCACGCCGTC CGCAGCGAGC GGCAGCGCGT
 31561 CCGTTCCGA CGCGATCACCG GCGGCCATCG CCCCCGCCGA CGGCAGCGCC TGCATCAGG
 45 31621 GGGCCCGTGC GGACACCAGC CTGACCGCGT CCTCCAGGGA CCAGACGCCG GCGACGTACG
 31681 CGCGGGCCAG CTCGCCGATC GAATGCCCA CGAAGGCCG CGGGCGTACG CCCCACGCC
 31741 CGAGCTGTGCG CGCGAGTGCG ACCTGGAGCG CGAACACCGC GGGCTGGCG TACCCGGTGT
 31801 CGTGGAGGTC GAGCCCGGCC GGCACGTCGA GGGCGTCCAG CACCTCGCGG CGAGTGCAGG
 31861 CGAAGACGTC GTAGGCGGCC GCGACGTCGT CCTCCAGGGA CGAGACGCCG GAGCCCTGTC
 31921 CGGAGAACAG CCACACGAGG CGGCGGTCCG GTTCTGCGGC CGCGGTGACC GTGTGGTGC
 50 31981 CGATCAGCGC GGGCGGTGCG GGGAAAGGCCG TGCGGGCGAG CAGGCCGCGC GCCACCGCG
 32041 GCTCGTCCTC CTCGCCGGTG GCGAGGTGGG CGCGCAGGCC GTGTACCTGT CGCTCGAGTG
 32101 CCTCGGGGT GCGTGCAGAG AGCAGCAGGG CGAGCGGTCC GGTGTCGGGT GCGGGGCCG
 32161 GTTCGGGGGC CGGTGCCGGG TGGCTTCGA GGATGATGTG AGCGTTGGTG CCGCTAACGC
 32221 CGAAGGGAGA CACCCCGGCC CGCGGTGGGG GTCTGGTTTC GGGCAGGGGG CGGGCGTCGG
 55 32281 TGAGGAGGTC GACGGCGCCG CGCGTCAGT CGACGTGCGA GGACGGCGTG TCCACGTGCA
 32341 GGGTGCAGGG CAGGGTGCCG TGCGCATGG CGAGGACCAT TTGATGACA CGGGCGACGC
 32401 CGCGGGCGGC CTGAGTGTGG CCGATGTTGG ACTTCAGCGA GCGCAGCGAG ACCGGGGTGT
 32461 CGCGATGCTG CCCGTAGGTG GCCAGTACCG CCTGCCCTC GATGGGGTCC CCCAGCCTGG
 32521 TCCCGGTGCC ATGCGCCTCG ACAGCGTCCA CATCGCCGG GGTGAGCCCG GCGTTGGCCA
 60 32581 GCGCCTGCCG GATCACCCGC TCCTGCGACG GCGCGTTCGG CGCCGACAAC CGCTTGGAAAG

	32641	CACCGTCCGT	GTTGACCAGCC	GAACCAACGCA	CGACCGCCAG	GACATTGTGG	CCGTGCCGCT
	32701	CGGCCTCGGA	GAGCCTCTCG	ACGATCAGCA	CACCGGATCC	CTCGGCAGAA	CCGGTGCCAT
5	32761	CAGCCGCATC	CGCGAACGCC	TTGCAGCGGC	CGTCCGGGGA	GAGGCCCCGC	TGCTGGGAGA
	32821	AGTCCACGAA	GCCGGACGGC	GAGGCCATCA	CCGTGACGCC	GCCGACCCACG	GCGAGCGAGC
	32881	ACTCCCCCGA	GCGCAGCGAC	TGCCCCGCC	GGTGCAGCGC	CACCAGCGAC	GACGAACACG
	32941	CCGTGTCCAC	CGTGACCGCC	GGACCCCTCCA	AACCGTAGAA	GTACGACAGC	CGACCCGACA
10	33001	GCACACTGGT	CTGGGTGCTG	GTGGCACCGA	AACCGCCGCG	GTGCGCTCCA	GTGCCGTACC
	33061	CGTAGAAGTA	GCCGCCCATG	AACACGCCGG	TGTCGCTTCC	GCGCAGCGAC	TCCGGGAGGA
	33121	TCCC GGCGTG	TTCCAGCGCC	TCCCACGAGG	TCTCCAGGAC	CAGACGCTGC	TGCGGGTCCA
15	33181	TCGCCAGCGC	CTCACGCGA	CTGATCCCAGA	AGAACGCCG	GTGCAAGTCC	GCCACCCCGG
	33241	CGAGGAAGCC	ACCATGACGC	ACGGTCGAAG	TGCCCGGATG	ATCCGGATCG	GGATCGTACA
	33301	GCCC GTCCAC	GTCCCCAACCA	CGGTCCTGCG	GAAACGCCGT	GATCCCCTCA	CCACCCGACT
	33361	CCAGCAGCCG	CCACAAGTCC	TCCGGCGACG	CGACCCACC	CGGCAGCCGG	CAGGCCATCC
	33421	CCACGATCGC	CAACGGCTCG	TCC TGCCCGA	CGGGCGCGGT	CGGGGTACGC	CGCCGGGTGG
20	33481	TGGCCCGCGC	GCCGGCCAGT	TCG TCCAGGT	GGGGCGCGAG	CGCCCTGCGC	GTGGGGTGGT
	33541	CGAAGACGAG	CGTAGCGGGC	AGCGTCAGGC	CGCTCGCGTC	GGCCAGCCGG	TTGCGCAGTT
	33601	CGACGCCGGT	CAGCGAGTCG	AAGCCCACCT	CCCTGAACGC	GCGCGCGGGT	GCGATGGCGT
	33661	GGGC GTCGCG	GTGGCCGAGC	ACCGCGGCAG	CGCTGGTACG	GACGAGGTGCG	AGCATGTCG
	33721	GCGCGGCCGG	AGGTGCGGAC	GTGCGCCGGA	CGGCCGGCAC	GAGGGTGC GT	AGGACCCGGC
25	33781	GGACCCGGTC	GGACGCGGCC	ACGGCGGCAG	GGTCA GAGCCG	GATCGGCACG	AGCGCGGGCC
	33841	GGTCGGTGTG	CAGGGCCGGC	TCGAACAGGG	CGAGCCCCCTG	TGCGGCCGTC	ATCGGGGTCA
	33901	TGCCGTGCG	GGCGATGCGG	GCCAGGTGCG	TGGCGGTGAG	CCGCCCCGCC	ATCCCGTCCG
	33961	CCGCGTCCCA	CAGTCCCCAG	GCGAGCGAGA	CGGCGGGCAG	CCCCTGGTGG	TGCCGGTGGC
	34021	GGGCGAGCGC	GTCGAGGAAC	GCGTTGCCGG	TCGCGTAGTT	GGCCTGACCC	GCGCCGCCGA
30	34081	ACGTGGCGGA	TATGGACGAG	TACAGGACGA	ACCGGCCAG	GTGCAAGATCG	CCGCGTACGCT
	34141	CGTGCAGGTG	CCAGGCAGC	TCCG CTTGA	CCCGCAGCAC	GGCGTCCCAC	TGCTCCGGCC
	34201	GCATGGTCGT	CACGGCCGCG	TCGTCGACGA	TCCC GGCCAT	GTGCA CGACG	GCGCGCAGCC
	34261	GCTGGCGAC	GTCGGCGACG	ACTGCGCCA	GCTCGTCCG	GTGCA CGACG	TCGGCGGCCA
	34321	CGTACCGCAC	GCGGTCTCG	TCCGGCGTGT	CGCCGGGCCG	GGCGTTGCGG	GACACCACGA
35	34381	CGACCTCGGC	GGCCTCGTGC	ACGGTGAGCA	GGTGGTCCAC	GAGGAGGC	CCGAGCCCGC
	34441	CGGTGCCGCC	GGTGACGAGG	ACGGTCCCAG	CGGTCAGCGG	GGAGGTTCCG	GTGGCCGCCG
	34501	CGACACGGCG	CAGACGGGCC	GCACCGCCTG	TGCGTCGGC	GACCCGGACG	TGCGGCTCGT
	34561	CGCCGGCGGC	GAGCCCGGCC	GCTATGGCGG	CGGGCGTGT	CTCGTCCGCT	TCGATCAGGG
	34621	CGACCGGCC	GGGATGCTCC	GTCTCCGCCG	TCCGGACCA	GCCGCCGAGC	GCTTCCCTGCG
40	34681	CGGGATCGCC	GGTACGGGTG	GCCACGATGA	GCCGGGATCG	CGCCCGAGCGC	GGCTCGGCAG
	34741	GCCAGGTCTG	CACGGTGGTG	AGCAGGTGCG	GGCCCCAGCTC	CGGGGTCCGG	GCGCCGGCG
	34801	AGGTGCCCGG	GTCGCGGGGT	TCCACCGGCA	GGACCCACGAC	CGGGGGGTGC	TCGCCGTGCG
	34861	GCACGTCCGC	GAGGTACGTC	CAGTCGGGGG	CGGGGTGACGC	GGGCACGGGC	ACCCAGCGA
	34921	TCTCGAACAG	CGCCTCGGCA	TCCGGGTGCG	CGGCCCCGAC	GGTCAGGCTG	TCGACGTCAA
45	34981	GGACCGGTGA	GCCGTGCTCG	TCCGTGGCGA	CGATCGGAC	CATGTGCGGG	CCGACCGCGT
	35041	CCAGCAGCAC	GCGCAGGCCG	GTCCGGCGCG	CGCGTGGAT	CCTCACGCCG	GACCAGGAGA
	35101	ACGCCAGCCG	GCGCCGCTCC	GGGTCCGTGA	AGACCGTCCC	GAGGGCGTGC	AGGGCCCGCT
	35161	CGAGCAGCAC	GGGGTGCAGC	CCGTACCGGG	CGTGGTGAG	CTGTTGGCG	AGGC GGACCG
	35221	ACCGTAGGC	GCGGCCCTCC	CCC GTCCACA	TCGCGGTAT	GGCCC GGAAAC	GCGGGCCCGT
50	35281	ACGAGAGCGG	CAGCGCGTCG	TAGAAGCCGG	TCAGGTCCGC	CGGGTCCGGC	TCGGGGGCGC
	35341	GCCAGTCCAC	GGGCTCCGCC	GGACCGCCAG	TGTCCACGCT	CAGCGCTCCG	GTCGCAC TGA
	35401	GCGCCCAGGG	GCCC GTGCCG	GTACGGCTGT	CGAGACTCAC	CGACCGCCGT	CCGGACACCT
	35461	CGGTTCCGAC	GGTGGCCTGG	ATCTCCGTGT	CGCCGTGCG	GTCGACCACC	ACCGGGCGA
	35521	CGATGGTCAG	CTCCCGCGATC	TCCGGCGTGC	CGAGCCGGGC	TCCCGCTTCG	GCGAGCAGTT
55	35581	CCACGAGCGC	CGAGCCGGGC	ACGATGACCC	GGCGTCCAC	CTCGTGGTCG	GCGAGCCAGG
	35641	GCTGACGGCG	TACCGAGACA	CCGCGGTGGC	CAGCGCGCC	TCGCCGTGCG	GCGAGGTGCA
	35701	CCCACGAGCC	GAGCAGCGGG	TGGCGGGACG	TTCCCGCCGG	TTCCCGTGTG	ATCCAGTAGC
	35761	GGTCACGGCG	GAACGGGTAC	GTGGGCAGCG	GCACCA CCCG	ACCGCGTGC	AACGACCA
	35821	TGACGGGCAC	GCCCCGGAC	CAGAGCGCGG	CGAGCGACCG	AGTGAAGCGG	TCCAGGGCGC
60	35881	CCTCGCCTCG	CCGCAGTGTG	CCGGTGCACG	CCGTATGCGC	ATGCCCGGGC	AGCGTGTCC
	35941	CCAGTGC GGT	GGTGAGCAGC	GGATGCGCGC	TGACCTCGAC	GAACCGCGG	TATCCGGGT
	36001	CCGCCAGGTG	GCCGGTGC	CGGGCGAACC	GAACGGTGC	GCGCAGGTTG	TCGTACCA
	36061	AGGCGGCGTC	CGCGGGCCGG	TCCAGCCACG	CCTCGTCCAC	GGTGGAGAAG	AACGGGACGT
	36121	CCGGCGTGC	CGGAGTGTATG	CCGGCGAGAG	CGTCGAGCAG	CGGCCGCCG	ATCGTTTCGA
	36181	CATGCGCGGT	GTGCGACGCC	TAGTCGACGG	CGATCCGGCG	GGCGCGGGGG	GTGGCGGCCA

	36241	CGAGCTCCTC	CACGGCGTCG	GCCGCACCGG	CGACAACGAT	CGACGCGGGT	CCGTTGACCG
	36301	CGCGCACCTC	CAGGCGCCCG	GCCCCACACGG	CGGCGTCGAA	GTCGGCGGGC	GGCACCGAGA
	36361	CCATGCCGCC	CTGCCCGGCC	AGTCGGTGG	CGACGAGTCG	GCTGCGCACC	GCGACGACCT
5	36421	TGCGGGCGTC	GTCCAGGGTG	AGCACCCCGG	CGACGCAGGC	CGCGCGACT	TCGCCCTGGG
	36481	AGTGGCCGAC	GACCGCGGCC	GGGGCGACCC	CGTGCACG	CCACAGCTCC	GCCAGCGCCA
	36541	CCATCACCGC	GAACGACCGC	GGCTGCACGA	CATCGACCCG	GTCGAACCGC	GGCGCTCCGG
	36601	GCCGCTGGGC	GATGACGTCC	AGCAGGTCCC	ATCCGGTGTG	CGGGGCGAGC	GCCGTGGCGC
	36661	ACTCGCGGAG	CCGCCGGGGC	AACACGGGCT	CGGTGGCGAG	CAGTTCGGCA	CCCATGCCGG
	36721	CCCACCTGGGA	GCCCTGCCCG	GGGAACCGGA	ACACGACACG	TGTGTCGGTG	ACGTCGGCGG
10	36781	TTCGGTCAC	GGCCCCCGGC	ACTTCGGCAC	CACGGCGAA	CGCCTCCGCC	TCTCGGGCGG
	36841	GCACGACCGC	CCGGTGGCGC	ATGGCCGTCC	GGGTGGTGGC	GAGCAGTGG	CCGACCGCGG
	36901	CCGCGCGGCC	AGTGAAGCGGG	GCCAGCTGTC	CCGCGACGTC	CCGAGTCCC	TCCGGGTCC
	36961	GGGCCGACAT	CGGCCAGACC	ACGTCCTCGG	GCACCGGCTC	GGCTTCGGGT	GCGGACACGG
15	37021	GTGCGGGCGC	GGCGGGGGGC	CCGGCCTCCA	GGACGACATG	GGCGTTGGTG	CCGCTGATGC
	37081	CGAACGACGA	GACACCCGCA	CGCCGGGCGC	GCCCGBTGAC	CGGCCACGGC	TCACTGCGGT
	37141	GCAGCAGCCG	GATGTCGCGC	TCCCAGTCGA	CGTGGGGGA	CGGCTCGTCC	ACGTGCAGCG
	37201	TGCGGGCAG	GACGCCGTG	CGCATCGCA	TGACCATCTT	GATGACGCCG	GCGACGCCGG
	37261	CCGCGGCCTG	GGTGTGGCCG	ATGTTGACT	TGAGCGAGCC	GATCAGCAGC	GGATGACCGC
20	37321	GTTCGCGCCC	GTAGGCCACT	TGCAAGGGCT	GGGCTCGAC	GGGGTCGCCG	AGACGGGTGC
	37381	CGGTGCCGTG	TGCCTCCACG	GCCTCGACGT	CACCCGGCGC	CAGGCCGGCG	TCGGCGAGCG
	37441	CACCGTGGAT	GACCGCGCTG	TGCGCAGGGC	CGTTGGGGC	GGACAGCCCC	TTCGACGCCG
	37501	CCTCGGAGTT	GACCGCGGGAG	CCGCGCACCA	GCGCCAGCAC	GGGGTGGCCG	TGGCGGGTGG
	37561	CCTCGGAGAG	CCGCTCCAGC	ACCAGGACAC	GGGCGCCCTC	GGCGAAGCTC	GTGCCGTCCG
25	37621	CGGTGTCCGC	GAAGGCCTG	GCACGGCCGT	GGGGGGCGAG	CCCGCGCTGC	CGGGAGAACT
	37681	CGACGAACCC	GGTCGTGTC	GCCATCACCG	TGACACCGCC	GACCAAGGGC	AGCGAGCACT
	37741	CCCCCGAGCG	CAGCGACCGC	GCGGCTGGT	GCAGCGCCAC	CAGCGACGAC	GAACACGCCG
	37801	TGTCGACGGT	GACCGACGGG	CCCTCCAGAC	CGAAGTAGTA	CGAGAGCCGC	CCGGAGAGAA
	37861	CGCTGGTCGG	CGTGGCGGTG	GCCCCGAAAC	CGCCCAGGTC	CACGCCCGCG	CCGTAGCCCT
	37921	GGGTGAACGC	CCCCATGAAT	ACGCCGGTGT	CGCTGCCCG	GACGTTTTCG	GGCAGGATGC
30	37981	CCGCTCGTTC	GAACGCCTCC	CACGACGTT	CGAGGACAG	ACGCTGCTGC	GGGTCCATCG
	38041	CCAGGCCCTC	ACCGGGCTG	ATCCCGAAGA	ACGCGCGTC	GAAGTCGGCG	GCGCCGGTGA
	38101	GGAAGCCGCC	GTGACGCACG	GAAACCTTGC	CGACCGCGTC	GGGGTTCGGG	TCGTAGAGCG
	38161	CGGCGAGGTC	CCAGCCGCCG	TCGGCGGGGA	ACTCGGTGAT	CGCGTCCCCG	CCGGAGTCGA
35	38221	CCAGCCGCCA	CAGGTCTTCC	GGTGACCGCA	CGCCACCGGG	CATCCGGCAC	GCCATGGCCA
	38281	C3ATCGCCAG	CGGCTCGTTC	CCCGCCACCG	TGGGTGCGGG	CACTGTCGCC	GCCGGAGCGG
	38341	CAGGGGCCGG	CTCACCCCGC	CGTTCTCAT	CCAGGCGGGC	GGCGAGCGCG	GCCGGTGTGC
	38401	GGTGGTCGAA	GACGGCCGTC	GCGGAGAGCC	GTACCCCGT	CGTCTCGGCC	AGGCTGTTGC
	38461	GCAACCGGAC	ACCGCTGAGC	GAGTCGATGC	CGAGGTCTT	GAACGCCGTC	GTGGCGTGA
40	38521	TCTCGGAGGC	GTGGCGTGG	CCGAGCACGG	CGGGCGTGGC	CGCACACACG	ATGGCCAGCA
	38581	GGTCACGATC	GGGGTCCGGG	TGCGGTGCG	GGTTGTCTC	CGCACGGCG	GCGATGCCGG
	38641	GCTCGGTCCG	CTGCCGGACG	GGCTCGGTGG	GAATEGCCG	GACCATGAAC	GGCACCGTCCG
	38701	CGGCGAGGCT	CCGCTCGATG	AAAGTGGGTG	CCTCGGCCCTC	GGTGAGCGGC	CGGAACCCGT
	38761	CGCGCACCCG	CTGCCGGTCG	GCCTCGTCAA	GTTGTCCGGT	GAGGGTGTG	GTGGTGTGCC
	38821	ACATGCCCCA	GGCGATGGAG	GTGGCGGGTT	GGCCGAGGGT	GTGGCGGGTGG	GTGGCGAGGG
45	38881	CGTCGAGGAA	GGCGTTGGCG	CGGGCGTAGT	TTCTTGTCC	GGGGCTGCCG	AGGACGGCGG
	38941	CGGCGCTGGA	GTAGAGGACG	AAAGTGGGTG	GGGGTTGGT	TTGGGTGAGG	TGGTGCAGGT
	39001	GCCAGGCCGC	GTGGCTTTG	GGGTGGAGGA	CGGTGGTGAG	GCGGTCGGGG	GTGAGGGCGT
	39061	CGAGGATGCC	GTCTCGAGG	GTGGCGGGCG	TGTGGAAGAC	GGCGGTGAGG	GGTTGGGGGA
	39121	TGTGGCGAG	GTGGTGGCG	AGTTGGTGGG	GGTCGCCGAC	GTCCGAGGGG	AGGTGGGTGC
50	39181	CGGGGGTGGT	GTGGGGGGT	GGGGTGCAGGG	AGAGGAGGT	GGTGTGGGGG	TGGTTCAGGT
	39241	GGCGGGCGAG	GATGCCGGCG	AGGGTGCAGGG	AGCCGCCGGT	GATGATGATG	GCGTGTTCGG
	39301	GGTGAGGGG	GGTGGTGGTG	GGTGGGGTGG	TGGTGTGGAG	GGGGGTGAGG	TGGGGTCGGT
	39361	GGAGGGTGTG	GTGGGTGAGG	CGGAGGTGGG	GGTGGTCAG	GGTGGCGAGT	TGGGCCAGGG
	39421	GGAGGGGAGT	GTGGGGGTGG	TCGGTTGCA	TGAGGCAGGAT	GGCGTGGGGG	TGTTCTGTTCT
55	39481	GGGGCGGTGCG	GGTGAGGCCG	GTGACGGTGG	CGCCGGCGGG	GTGGTGGGGT	GTGTGGACGA
	39541	TGAGGGTGTG	GTCGGTGGTG	GTGAGGTGGT	TGTCAGGGC	GGTCAGGACG	CGGGTGGCGC
	39601	GGGTGTGGGC	GGGGGTGGGT	ATGTCCTCGG	GGTCGTGCCG	GTGGCGGGCG	GTGATCAGGA
	39661	C3TGTCCCTC	GGGCAGGTCA	CCGTCGTAGA	CCGCTCGGC	GACCGCGAGC	CACTCCAAACC
	39721	GGAGCGGGTT	CGGCCCCGAC	GGGGTGTGCG	CCCGCTCCCT	CAGCACCAGC	GAGTCCACCG
60	39781	ACACGACAGG	ACGGCCATCC	GGGTGCGCCA	CGCGCACGGC	GACGCCGGCC	TCCCCCGGGG

39841 TGAGGGCGAC GCGCACCGCG GCGGCCCGG TGGCGTTCA GCGCACGCC STCCAGGAGA
 39901 ACGGCAGCTC GATCCCGCCG CCCGCCTGA GGCGCCCGC GTGCAGGGCC GCGTCGAGCA
 39961 GTGCCGGATG CACACCGAAA CGTCCGCT CGGCCTG CTCGTCGGG AGCGCCACCT
 40021 CGGCATACAC GGTGTACCA TCACGCCAGG CAGCCCGAA CCCCTGGAAC CGCACCCGT
 5 40081 ACTCATAACCG GGCATCCCG AGTCGTCAT AGAACCCGA GACGTCGACG CGCGCCGGCG
 40141 TGGCCGGCGG CCACTGCGAG AACGGCTCAC CGGAAGCGTT GGAGGTATCC GGGGTGTCGG
 40201 GGGTCAGGGT GCCGCTGGCG TGCCGGGTCC AGCTGCCGT GCCCTCGGTA CGCGCGTGG
 40261 CGGTACCCGG CGGCCGTCCG GCCTCATCGG CCCCTCCAC GGTACCCGAC ACATCCACCG
 40321 CTGCGGTAC CCGCACCCAG AGCGGGGATT CGATGACCAG TTCATCCACC ACCCGCAAC
 10 40381 CGGTCTCGTC ACCGGCCCGG ATGACCAAGCT CCACAAACGC CGTACCCGGC AGCAGAACCG
 40441 TGCCCGCAC CGCGTGATCA GCCAGCCAGG GATGCGTACG CAATGAGATC CGGCCGGTGA
 40501 GACAACACCC ACCACCGTCC TCGGGGGCA GTGCTGTGAC GCGGGCCAGC ATCGGATGCG
 40561 CGGCCCCGGT CAGCCCGGGC GCGGACAGGT CGGTGGCACC GGGCGCTCC AGCCAGTACC
 40621 GCCTGTGCTC GAACGCGTAG GTGGGCAGAT CCAGCAGCCG CCCCAGGACCC GTTCGACCA
 15 40681 CGTGGCCCA GTCCACCCCCC GCACCCAGAG TCCACGCTG CGCCAACGCC CCCAGGCCAC
 40741 GCTCCCAGCC ACCGTACCA GTCCGCAACG ACGCCACCGT GCGGGCCTGT TCCATCGCCG
 40801 GCAGCAGCAC CGGATGGCA CTGCACTCCA CGAACCCGA CCCGTCAGC TCCGCCACCG
 40861 CGGCATCCAG CGCGACAGGG CGACCGAGGT TCCGGTACCA GTACCCCTCA TCCACCGGCT
 40921 CGGTACCCCA GGCCTGTCC ACGGTCGACC ACCACGCCAC CGACCCGGTC CGGCCGGAAA
 20 40981 TCCCCCTTCAG TACCTCAGCG AGTCGTCCT CGATGGCCTC CACGTGAGGC GTGTGGGAGG
 41041 CGTAGTCGAC CGCGATAACGA CGCACCCGCA CCCCACATCAGC CTCATACCGC GCCACCCACT
 41101 CCTCCACCGC CGACGGGTCC CCCGCCACCA CGTCTGAAAGC CGGACCATTA CGCGCCGCGA
 41161 TCCACACACC CTCGACCAAGA CCCACCTCAC CGGCCGGCAA CGCCACCGAA GCCATCGCCC
 41221 CCCGGCCGGC CAGCCGCGCC GCGATCACCC GACTGCGCAA CGCCACCCAGC CGGGCCGGGT
 25 41281 CCTCCAGGCT GAGGGCTCG GCCACACACG CGGCCGCGAT CTCCCCCTGC GAGTGTCCGA
 41341 CCACAGCGTC CGGCACCGAC CCATGCGCCT GCCACAGCGC GGGCAGGCTC ACCGCGACCG
 41401 CCCAGCTGGC CGGCTGGACC ACCTCCACCC GCTCCGCCAC ATCCGACCGC GACAACATCT
 41461 CCCGCACATC CCAGCCCGTG TCGGCAACAA ACGCCCGCGC ACACCTCTCC ATACGAGCG
 41521 CGAACACCCG GGAACGGTCC ATGAGTTCCA CGCCCATGCC CACCCACTGG GCACCCCTGCC
 30 41581 CGGGGAAGAC GAACACCGTA CGCGCTGTAT CCACCGCCAC ACCCATCAC CGGGCATCAC
 41641 CCAGCAGCAC CGCACGGTGA CGGAAGACAG CACGCTCACG CACCAACCCC TGCGCCACCG
 41701 CGGCCACATC CACCCCCACCC CGCGCGAGAT ACCCCTCCAG CGCTCCACC TGCCCCCGCA
 41761 GACTCACCTC ACCACGAGCC GACACGGCA CGGGCACCAA CCCATCACCA CCCGACTTCA
 41821 CACGCCACGG CCCAGGAACA CCCTCCAGGA TCACGTGCGC GTTCGTACCG CTCACCCCGA
 35 41881 ACGACGACAC ACCCGCATGC GGTGCCCGAT CCGACTCGGG CCACGGCCTC GCCTCGGTGA
 41941 GCAGCTCCAC CGCACGGGC GACCAGTCCA CATGCGACGA CGGCTCGTCC ACGTGAGCG
 42001 TCTTCGGCGC GATCCCATGC CGCATCGCA TGACCATCTT GATGACACCG GCGACACCCG
 42061 CAGCCGCGCTG CGCATGACCG ATGTTGACT TGACCGAACCG GAGGTAGAGC GGCGTGTCGC
 42121 GGTCTGCCCG TAGGCCCGCG AGGACGGCT CGCCTCGGAT CGGGCTGCC AGCCGCGTGC
 40 42181 CGGTGCCGTG CGCCTCCACC ACGTCCACAT CGCGGGCGCG CAGTCCGGCG TTGACCAAACG
 42241 CCTGCCGGAT CACGCGCTGC TGGCGACGC CGTGGGGGGC GGACAGTCCG TTGGAGGCAC
 42301 CGTCCTGGTT CACCGCCGAG CGCGGGACGA CGCGGAGAAC GGTGTGCCCG TTGCGCTCGG
 42361 CGTCGGAGAG CGCTCCACG ACGAGAACGC CGACGCCCTC GCGGAAGCCG GTCCCGTCCG
 42421 CGCGTCCGGC GAACGCCCTG CACCGTCCGT CGGGGGAGAG TCCGCGCTGC CGGGAGAACT
 45 42481 CCACGAGCTC TCGGGTGTTC GCCATGACGG TGACACCGCC GACCAGCGCC AGGGAGCACT
 42541 CCCCGCCCG CAGTGCCTGT GCCGCCCTGGT GCAGGGCGAC CAGCGACGAC GAGCACGCCG
 42601 TGTCGACCGT GACCGCCGGG CCCTGAAGTC CGTACACGTA CGAGAGGGCGC CGGGACAGGA
 42661 CGCTCGCTG CGTCGCCGTG ACACCGAGCC CGCCCAAGGTC CGGGCCGACG CCGTAGCCCT
 42721 GGTTAACGC GCCCATGAAC ACGCCGGTGT CGCTCTCCCG GAGCCTGTCC GGCACGATGC
 50 42781 CGGCCTCTC GAACGCCCTC CAGGAGGTCT CGAGGATCAG CGCCTGCTGG GGGTCCATCG
 42841 CCAGCGCCCTC GTTCGGACTG ATGCCGAAGA ACAGCGCGTC GAACCCGGCG CGGGCCAGGA
 42901 ATCCGCCGTG GCGTGTGCGT GAGCGGCCGG CGCGTCCGG GTCCGGGTG TACAGCGCGT
 42961 CGACGTCCCA GCCCCGGTCG GTGGGGAACT CGGTGATCGC CTCGGTACCG GCGGCACGA
 43021 GCGCCACAG GTCCCTCCGGC GAGGCACCC CGCCGGGAG TCGGCACGCC ATGCCGACGA
 55 43081 CGCGACGGG GTCGCCGGAG CCGAGGGTCT GGGCGGTGCG GGGTGCCT GTCGCGGAGC
 43141 CGGCAGGGTG GGCAGCGAAC GCACCGGGAG TGGGGTGGTC GAACCGGGTT GACGCAGGGCA
 43201 CGCGCAGACC CGTCCGCGCG GCGAACGGTGT TGGTAAACTC GACGGTGGTG AGCGAGTCGA
 43261 GCGCGTCTC GCGGAACGTG CGGTCCGGGG AGCAGTGTCC GGCGCCCGGG AGGCCAGGA
 43321 CGGTGGCGAC GCTGTGCGGG ACCAGGTGCA GCAGTACGTC CTCCCGGCCCG GACACGGCCG
 60 43381 CGCGAGGGCG GTTCGCCAC TCCTGTTCCG TGGCGTCCGG CTCGGCCGGT CCGGTCACTG

43441 CGGTGAGGAT CGGGGGCGTG GCGCCCGCCA TCGTCGCCGC CCGCGCCCCG GCGGAACCGG
 43501 TCCGGGCCAC GATGTACGAG CGCCCGCCCC CGATGGCCTT CTCGATCAGG TCGCCGGTGA
 43561 GCGCCGGCGG TTGATGCCG GGCAGCGCGC GGACGGTGAC GGTGGGAGT CCCTCCGCGG
 43621 CCCGTGGCGG GGTGTGGCG TCGGCGCCGG CCGGGCCGTC GAGCAGGACG TGCACGAGCG
 5 43681 CGCCGGGGTT CGCGGCTTCC TCGGCTGCCG TGGTCACGTG GGTGAGGCCG GTCTCGTCG
 43741 GGAGCAGGCC GGCAGCGGTG TCGGCGTCTT CCCCAGGTGAC CAGGACCGGC GCGTCGGGGC
 43801 CGATCGGAGG CGGCACGGTG AGGACCATCT TGCCGGTGTG CCGGGCGTGG CTCATCCACG
 43861 CGAACCGGTC CCGCGCACGG CGGATGTCCC ACGGCTGCAC CGGCAGCGGG CACAGCTCAC
 43921 CGCGGTGCAA CAGGTGAGG AGCAAGTCGA GGATCTCCC CAGGCGCGGC GGATCCACGT
 10 43981 CGGCCAGGTC GAACGGCTGC TGGGGCGCGT GGCGGATGTC GGTCTTGCCT ATCTCGACGA
 44041 ACCGGCCGCC CGGTGCGAGC AGGCCGATGG ACGCGTCGAG GAGTTCACCG GTGAGCGAGT
 44101 TGAGCACGAC GTCGACCAGG GGGAAAGGTGT CGGCGAACGC GGCCTGCAGG GAGTTCGCCA
 44161 CATGGTCGGT GTCGAAGCCG TCGGCGTGCAG CAGGTGTTG TTTGGGGGA CTGGCGGTGG
 15 44221 CGTACACCTC GGCGCCGAGG TGGGGCGCA TCGGGTCGC CGCCATGCCG ACACCCCGC
 44281 TCGCGCGTG GACCAAGGACC TTCTGGCGG GTCGCAGCTC GCCCAGCTCG ACGAGGCCGT
 44341 ACCAGGCAGT GGCAGAACACG ATGGGCACGG ACGCGCGAT GGGGAACGAC CATCCCCGTG
 44401 GGATCCGTGC GACCAAGCCG CGGTCGCCGA CCACGCTGC CGGAAACGCG TCCCTGCACGA
 44461 GACCGAACAC GCGGTCGCCG GGGGCCAGGT CGTCGACGCC GGGTCCGACT TCGGTACGA
 20 44521 TGCCCGCGGC CTCCCCGCC ATCTCGCCCT CGCCCGGGTA GGTGCGAGC GCGATCAGCA
 44581 CGTCGCGGAA GTTCAGCCCC GCGGCGCGG CGTCGATGCG GACCTCGCCG GCGGCCAGGG
 44641 GCGCGCGGG ACAGTCGAGCG GGGCGACGAC GAGGTGCGG AGCGTTCCGG AGGCGGGCGG
 44701 GCGCAGCGCC CACTGGCGG GTCGGCAGGG GGGTGGTGT CGCGCGTACCG AGCCGGGGCA
 44761 CGTAGGCCAC GCGGGCCCGC AGCGCGATCT GGGGTTCGCC GAGCGAGGCC GCGGCAGGG
 44821 CGAGGTCGTC ATCGCCGTCC GTGTCACCA GCACGAACGA TCGGGTTCG GCGGCCTGGC
 25 44881 GGCAGCGC CTCGTCCCAG AGCGGGCCT GGTCCGCGC CGGGATCTCG GCCGGGCCGA
 44941 CGCCCACCGC GCGGCGGGTG ACGACCGTCC GCGGGGTGA CGGGGTGCCG GGCAGGTGCG
 45001 GCGCTCCCA GACCAAGTTCG CACAGCGTGG CCTCGCAACT GCGGGTGGCG ACCAGATGGG
 45061 CCGGCAGCCC CGCGAGCCGC GCGCCTGGG CCTTGCCGA CGGGTGCAGG GGGATCGTGG
 45121 TGACGTGCCA GATCTCGTCG GGCACCTTGA AGTAGGCGAG CGGGCGGGCG CACTCGCGA
 30 45181 GGATCGCTC GGCAGGGACG CGGGGGCGT CGGAAACGAC GTAGAGCAGC GGTATGTCGC
 45241 CGAGGACGGG GTGCGGGCGG CCCGCCGC CGCGTCCCG GACACGGGCC ACCTCCTGGG
 45301 CGACGGTCTC GATCTCCCGG GGGTGGATGT TCTCCCGCC GCGGATGATC AGCTCCTTGA
 45361 CCCGGCCGGT GATCGTCACG TGTCCGGTCT CGGCCTGACG TCGGAGGTCC CGGGTGCAGG
 45421 ACCAGCCGTC CACGAGCACC TGGGCGTCG CCTCCGGCTG GCGTGGTAG CCGAGCATGA
 35 45481 GGCTCGGCC GCTCGCCAC AGCTCGCCCT CCTCGCCGGG TGCCACGTG GCGCCGACA
 45541 CGGGTGCAC GACCGCAGC GACAGGCCG GCACGGGAG CCCGACAGAG CGGGAAACCC
 45601 GCGCATCCCTC CAGGGTGTG GCGGTGAGCG AGCGGTCGT CTCGGTGCAG CGTACGTGT
 45661 CGAGCAGGGG CACGCCGAAC GTGCCCTCGA AATCCCTGGT GAGCGACGCC GGCGAGGTGG
 45721 ATCCGGCGAC CAGCGCCACG CGCAGCGCG GAGCCCGCG CTGGCCGGAC ACGGCGCCGA
 40 45781 GGAGGTAGCG GTACATCGTC GGCACGCCGA CGAGCACGGT GCTGGAGTGT TCGGCCAGGG
 45841 CGTCGAGGAC GTCACCGCGC ACGAAGCCGC CCAGGATAACG GCGGGACGCC CCGACCGTGA
 45901 GGACGGCGAG CAGGCAGAGG TGGTGGCCGA GCGTGTGAA CAGCGGGGCCG GGCAGAGCA
 45961 GTTCGTCGTC CTCGGTCAGC CGCCAGGACG GCACGTCGCA GTGCATCGCG GACCACAGGC
 46021 CGCTGCGCTG TGCGGAAACC ACGCCCTTGG GACGGCCGGT GGTGCCGGAG GTGTAGAGCA
 46081 TCCAGGGGGG TTGTCGCCAGG CCGAGGTCGT CGCGGGGCCG GCACGGCGGC TCGGTCCCGG
 46141 CGAGGTCTC GTAGGAGACG CAGTCGGTG CGCGGGGCCG GACCGAGCAGC ACGGTGGCGT
 46201 CGGTGCCGGT GCGGCGCACC TGTCGAGGT GGGTTTCGTC GGTGACCGAGC ACGGTGGCGC
 46261 CGGAGTCGGT CAGGAAGTGG GCGAGTTCGG CGTCGGCGGC GTCCGGGTTG AGCGGGACGG
 46321 CGACGGCGGC GGGCGGGCG GCGGGAGGT AGACCTCGAT GGTCTGATC CGGTTGCCGA
 50 46381 GCACCATCGC GACCCGGTCG CGCGGGTCGA CGCCGGACGC GGCAGGGTGT CGGGCGAGCC
 46441 GGCGGGCCCG GAGCGGAGT TCGTGTACG TCACGGCGCG TTGGAATCC GTGTAGGCAGA
 46501 TCCGGTCGCC GCGTCGCTCG GCATGGATGC GAGCAATTG GTGCAACGCC CGGATTGGTT
 46561 CCACACCGC CATGGAAACA CCTTTCTCTC GACCAACCGC ACAACAGCAC GGAACCGGCC
 46621 ACGAGTAGAC GCGGGCGACG CTAGGAGCGT TTCCGGACCC GCCACCCCT GAAGATCCCC
 55 46681 CTACCGTGGC CGGCCTCCCC GGACGCTCAT CTAGGGGGTT GCACGCATAC CGCCGTGCGT
 46741 AATTGCCCTC CTGATGACCG ATGCCGGACG CCAGGGAAAGG GTGGAGGCAGT TGTCCATATC
 46801 TGTCACGGCG CCGTATTGCCC GCTTCGAGAA GACCGGATCA CGGACCTCG AGGGTGACGA
 46861 GACGGTGCTC GGCCTGATCG AGCACGGCAC CGGCCACACC GACGTGTCGC TGGTGGACGG
 46921 TGCTCCCCGG ACCGCCGTGC ACACCAAGCAC CGTGACGAC GAGGGCTTCAG CCGAGGTCTG
 60 46981 GCACGCACAG CGCCCTGTGC AGTCCGGCAT GGACAACGGC ATCGCCTGGG CGCGCACCGA

	47041	CGCGTACCTG	TTCGGTGTG	TGCCGACCGG	CGAGAGCGGC	AGGTACGCCG	ATGCCACCGC
	47101	GGCCCTCTAC	ACGAACGTCT	TCCAGCTCAC	CCGGTCGCTG	GGGTATCCCC	TGCTCGCCCG
5	47161	GACCT3GAAC	TACGTCAAGC	GTATCAACAC	GACGAACGCG	GACGGGCTGG	AGGTGTACCG
	47221	GGACTTCTGC	GTGGGGCGCG	CCCAGGCCT	CGACGAGGGC	GGGATCGACC	CGGCCACCAT
	47281	GCCCCGGGCC	ACCGGTATCG	GCAGCCCACGG	GGGGCGGCATC	ACCTGCGTGT	TCCTCGCCGC
	47341	CCGGGGCGGA	GTGCGGATCA	ACATCGAGAA	CCCCGCCGTC	CTCACGGCCC	ACCACTACCC
10	47401	GACGACGTAC	GGTCCGCGGC	CCCCGGTCTT	CGCACGGGCC	ACCTGGCTGG	GCCCGCCGGA
	47461	GGGGGGCGCG	CTGTTCATCT	CCCGCAGGCC	CGGCATCCTC	GGACACCAGA	CGGTGCACCA
	47521	CGGTGATGTG	ACCGGCCAGT	GCAGGAGTCGC	CCTCGACAAC	ATGGGGCGGG	TCATCGGCGC
15	47581	GGAGAACCTG	CGGCGCCACG	GCCTCCAGCG	GGGGCACGTC	CTCGCCGACG	TGGACCCACCT
	47641	CAAGGTCTAC	GTCCGCGCC	GCGAGGATCT	CGATACTGGTC	CGCCGGGGTCT	GCGCCGCACG
	47701	CCTGTCGAGC	ACCGCGCCCG	TCGCCCTTT	GCACACCGAC	ATAGCCCCG	AGGATCTGCT
	47761	CGTCGAAATC	GAAGGCATGG	TGGCGTGACA	ATACCCGGTA	AAAGGCCCCG	GACGCTGCAC
20	47821	CTCGGGGAT	CCGCGAAGAG	AAAGAAGAGC	GTCACCGCAC	AGCGCGGCAG	CCCGGTCCTT
	47881	TCGTCTTCG	CACAGCGGC	GATCTGGTT	CTCCAGCAAT	TGACCCCGGA	GAGCAACGCC
	47941	TATAATCTCC	CGCTCGTCA	ACGCCCTGCGC	GGTCTATTGG	ACCGCGCCGC	CCTGGAGCGT
	48001	GCGCTGGCGC	TCGTCGTCG	GCGCCACGAG	GCGTTGCGGA	CGGTGTTCGA	CACCGCCGAC
	48061	GGCGAGCCCC	TCCAGCGGGT	GCTTCCCGCC	CCGGAACACC	TCCCTGCGCCA	CGCGCGGGCG
25	48121	GGCAGCGAGG	AGGACGCCGC	CCGGCTCGTC	CGCGACGAGA	TCGCCGCGCC	GTTCGACCTC
	48181	GCCACCGGGC	CGTTGATCAG	GGCCCTGCTG	ATCCGCTCTG	GTGACGACGA	CCACGTTCTC
	48241	GCGGTGACCG	TGCAACATGT	CGCCGGCGAC	GGCTGGTCGT	TCGGGCTCCCT	CCAACATGAA
	48301	CTCGCAGCCC	ACTACACCGC	GCTGCGCGAC	ACTGCCCCGC	CTGCGGAACCT	GCCGCGCGTTG
	48361	CCGGTGCAGT	ACGCCGACTT	CGCCGCCTGG	GAGCGGGCGC	AACTCACCAG	CGCCGGACTG
	48421	GACAGGCCTC	TGGCCTACTG	GCAGGAGCAA	CTCCGGGGCG	CCCCGGCGCG	GCTCGCCCTC
30	48481	CCCACCGACC	GTCCCCGCC	GCCGGTCGCC	GACGCGGACG	CGGGCATGGC	CGAGTGGCGG
	48541	CCGCGGGCCG	CGCTGGCCAC	CGCGGTCCCTC	ACGCTCCGCG	GCGACTCCGG	TGCGTCCGTG
	48601	TTCATGACCC	TGCTGGCGGC	CTTCCAAGCG	GTCCTCGCCC	GGCAGGCGGG	CACGCGGGAC
	48661	GTGCTGGTCG	GCACGCCGT	GGCGAACCGT	ACGCGGGCGG	CGTACGAGGG	CCTGATCGGC
	48721	ATGTCGTCA	ACACGCTCGC	GCTGCGCGGC	GACCTCTCGG	GCGATCCGTC	GTTCGGGAA
35	48781	CTCCTCGACC	GCTGCCGGGC	CACGACACG	GACGCGTTCG	CCCACGCCGA	CCTGCCGTT
	48841	GAGAACGTCA	TCGAACTCGT	CGCACCGGAA	CGCGACCTGT	CGGTCAACCC	GGTCGTCCAG
	48901	GTGCTGTTGC	AGGTGCTGCG	GCGCGACGCG	GCGACGGCCG	CGCTGCCCGG	CATCGCGGCC
	48961	GAACCGTTCC	GCACCGGACG	CTGGTTCACC	CGCTTCGACC	TCGAATTCCA	TGTGTACGAG
	49021	GAGCGGGGTG	GGCGCGCTGAC	CGGCGAACTG	CTCTACAGCC	GTGCGCTGTT	CGACGAGCCA
40	49081	CGGATCACCG	GGTTGCTGGA	GGAGTTCACG	GGGGTGCCTTC	AGGCGGTAC	CGCCGACCCG
	49141	GACGTACGGC	TGTCGCGGCT	GCCGGCCGGC	GACGCGACGG	CGGCAGCGCC	CGTGGTGC
	49201	TCGAACGACA	CGGCGCGGG	CCTGCCGTC	GACACGCTGC	CGGGCCTGCT	GGCCCCGGTAC
	49261	GCCGACCGCA	CCCCCGGGC	CGTGGCCGTC	ACCGACCCGC	ACATCTCCCT	CACCTACCGC
	49321	CAGCTGGACC	GGCGGGCGAA	CCGCCCTCGCG	CACCTGCTCC	GCGCGCGCG	CACCGCCACC
45	49381	GGCGACCTGG	TCGGGATCTG	CGCCGATCGC	GGCGCCGACC	TGATCGTCGG	CATCGTGGGG
	49441	ATCCTCAAGG	CGGGCGCCGC	TTATGTGGCG	CTGGACCCCG	AACATCCTCC	GGAGCGCACG
	49501	GCGTTGTCG	TGGCGACGC	GCAGCTGACC	ACGGTGGTGG	CGCACGAGGT	CTACCGTTCC
	49561	CGGTTCCCCG	ATGTGCGCGA	CGTGGTGGCG	TTGGACGACC	CGGAGCTGGA	CGGGCAGCG
	49621	GACGACACCG	CGCGCGGACGT	CGAGCTGGAC	CGGGACAGCC	TCGCCCTACGC	GATCTACACG
50	49681	TCCGGTCGA	CGCGCAGGCC	GAAGGCCGTG	CTCATGCCG	GTGTCAGCGC	CGTCAACCTG
	49741	CTGCTCTGGC	AGGAGCGCAC	GATGGGCCGC	GAGCGGGCCA	GCCGCACCGT	CCAGTTGTCG
	49801	ACGCCCCACGT	TCGACTACTC	GGTGCAGGAG	ATCTTTTCCG	CGCTGCTGGG	CGGCACGCTC
	49861	GTCATCCCGC	CGGACGAGGT	CGGGTTCGAC	CCGCCGGGAC	TCGCCCGGTG	GATGGACGA
	49921	CAGGGATTAA	CCCGGATCTA	CGCGCCGACG	CGCGTACTGC	CGCGCCTGAT	CGAGCACGTC
55	49981	GATCCGCACA	GCGACCAAGT	CGCCGCCCTG	CGGCACCTGT	GCCAGGGCGG	CGAGGGCGCTG
	50041	ATCCTCGACG	CGCGGTTGCG	CGAGCTGTGC	CGGCACCCGGC	CCCACCTGCG	CGTGACAAAT
	50101	CACTACGGTC	CGGCCGAAAG	CCAGCTCATC	ACCGGGTACA	CGCTGCCCGC	CGACCCCGAC
	50161	GCGTGGCCCG	CCACCGCACC	GATGGGCCCG	CCGATCGACA	ACACCCGCAT	CCATCTGCTC
	50221	GACGAGGCAGA	TGCGGCCGGT	TCCGGACGGT	ATGCCGGGCG	AGCTCTGCGT	CGCCGGCGTC
55	50281	GGCCTCGCCC	GTGGGTACCT	GGCCCGTCCC	GAGCTGACCG	CCGAGCGCTG	GGTGCCGGGA
	50341	GATGCGGTG	GCGAGGAGCG	CATGTACCTC	ACCGGCACCC	TGGCCCGCCG	CGCGCCCGAC
	50401	GGCGACCTGG	AATTCTCGG	CCGGATCGAC	GACCAGGTCA	AGATCCGCGG	CATCCCGCTC
	50461	GAACCGGGTG	AGATCGAGAG	CCTGCTGCC	GAGGACGCC	GCGTCACGCA	GGCGCGGGTG
	50521	TCCGTGCGCG	AGGACCGGGC	GGGGCAGAGAAG	TTCCCTGGCCG	CGTACGTCGT	ACCGGTGGCC
60	50581	GGCGGGCACG	GCGACGACTT	CGCCGCGTCG	CTGCGCGCGG	GACTGGCCGC	CCGGCTGCC

50641 GCGCGGCTCG TGCCCTCCGC CGTCGTCTG GTGGAGCGAC TGCCGAGGAC CACGAGCGGC
 50701 AAGGTGGACC GGCAGCGCGT GCCCGACCCG GAGCCGGGCC CGCGCTCGAC CGGGGCGGTT
 50761 ACGCCCCGCA CCGATGCCGA CGGGACGGTG TGCCGGATCT TCCAGGAGGT GCTCGACGTC
 50821 CGCGGGTCG GTGCCGACGA CGACTTCTTG ACGCTCGGCG GGCACCTCCCT GCTGCCACC
 5 50881 CGGGTCGTCT CCCGCATCCG CGCCGACCGT GGTGCCGATG TCCCCTGCG TAGCCTCTTC
 50941 GACGGGCGGA CGCCCCGCCGC GCTCGCCCGT GCGGCGGACG AGGCCGGCCC GGCCGCCCTG
 51001 CCCCCGATCG CGCCCTCCGC GGAGAACCGG CGGGCCCCCC TCACCGCGGC ACAGGAACAG
 51061 ATGCTGCACT CGCACGGCTC GCTGCTGCC GCGCCCTCCT ACACGGTCGC CCCGTACGGG
 51121 TTCCGGCTGC CGGGGCCACT CGACCGCGAA GCGCTCGACG CGGCACGTGAC CGGGATCGCC
 10 51181 GCGCGCCACG AGCCGCTGCG GACCGGGTTC CGCGATCGGG AACAGGTCGT CGGGCCGCC
 51241 GCTCCGGTGC GCGCCGAGGT GGTTCCGGTG CGGGTCCGGCG ACGTCGACGC CGCGGTCGG
 51301 GTCGCCACC GGGAGCTGAC CGGGCCGTTG GACCTCGTGA ACCGGTCGTT GCTGCGTGCC
 51361 GTGCTGCTGC CGCTGGGCGC CGAGGATCAC GTGCTGCTGC TGATGCTGCA CCACCTCGCC
 51421 GGTGACGGAT GGTCTTCGA CCTCTCTGGT CCGGGAGTTG CGGGGAGCAGA ACCGGACCTT
 15 51481 CCGGTGTCCT ACACGGACGT GGCCCGGTGG GAACGGAGTC CGGGCGTGAT CGGGCCAGG
 51541 GAGAACGACC GGGCCTACTG CGGCCGGCGG CTGGGGGCGC CAACCGCGCC GGAGCTGCC
 51601 GCGGTCCGGC CGGGCGGGGC ACCGACCGGG CGGGCGTTCC TGTGGACGCT CAAGGACACC
 51661 GCCGTCCTGG CGGCACGCCG GGTCCGGAC GCCCCACGAC CGACGTTGCA CGAAACCGTG
 51721 CTCGGCGCCT TCGCCCTGGT CGTGGCGGAG ACCGCGACCA CCGACGACGT GCTCGTCGCG
 20 51781 ACGCCGTTCG CGGACCGGGG GTACGCCGG ACCGACCAAC TCATCGGCTT CTTCGCGAAG
 51841 STCCTCGCGC TCGCCCTCGA CCTCGCGCGC ACGCCGTCGT TCCCCGAGGT GCTGCGCCGG
 51901 STGCACACCG CGATGGTGGG CGCGCACCCC CACCAGGGCGG TGCCCTACTC CGCGCTGCGC
 51961 GCCGAGGACC CCGCGCTGCC GCCGGCCCCC GTGCTGTTCC AGCTCATCAG CGCGCTCAGC
 52021 CGGAACTGC GGCTGCCCGG CATGACACCC GAGCCGTTCC CGTCGTCGCG CGAGACCGTC
 25 52081 GACGAGATGA CGGGCGAAGT GTCGATCAAC CTCTTCGACG ACGTCGCGAC CGTCTCCGGC
 52141 GCGGTGGTCC ACGATGCCGC GCTGCTCGAC CGTGCACCC CGTGCACCG TCGACGATT GCTCACCGG
 52201 STGGAGGCGA CGCTCGTGC CGCCCGGGGC GACCTCACCG TACCGCTCAC CGGTTACGTG
 52261 GAAAGCGAGT AGCCATGCC GAGCAGGACA AGACAGTCGA GTACCTTCGC TGGCGACCG
 52321 CGGAACTCCA GAAGACCCGT CGGAAACTCG CGCGCACAG CGAGCCGTTG GCGATCGTGG
 30 52381 GGATGGCTG CGGGCTGCCG GGCGGGTGC CGTCGCCGA GGACCTGTGG CAGTTGCTGG
 52441 AGTCCGGTGG CGACGGCATT ACCGCGTTCC CCACGGACCG GGGCTGGGAG ACCACCGCCG
 52501 ACGGTCCGG CGGCTTCCTC ACCGGGGCGG CGGCTTCGA CGCGCGTTTC TTCGGCATCA
 52561 GCCCGCGCGA GGCCTGGCG ATGGACCCGC AGCAGCGCCT GGCCCTGGAG ACCTCGTGGG
 52621 AGGCGTTCGA GCAACGCCGC ATCGATCCGC AGACGCTGCG GGGCAGTGAC ACGGGGTGT
 35 52681 TCCCTGGCGC GTTCTTCAG GGGTACGGCA TCGGCCCGA CTTCGACGGT TACGGCACCA
 52741 CGAGCATTCA CACGAGCGTG CTCTCCGGCC GCCTCGCGTA CTTCTACGGT CTGGAGGGTC
 52801 CGCGGTACACG GGTGACACCG CGTGTTCGT CGTCGCTGGT GGCCTGAC CAGGCCGGC
 52861 AGTCGCTGCG CTCCGGCGA TGCTCGCTCG CCCTGGTCGG CGCGTCACG GTGATGCC
 40 52921 CGCCGGCGG GTTCGCGGAC TTCTCCAGC AGGGCGGCCT GGCCTCGAAC GCGCGCTGCA
 52981 AGGCCTTCGC GGAAGCGGGCT GACGGCACCC GTTTCGCGGA GGGGTCCGGC GTCCCTGATCG
 53041 TCGAGAAGT CTCCGACGCC GAGCGCACCG GCCACCGCGT GCTGGCGGTC GTCCGGGTT
 53101 CCGCCGCAA CGAGGACGGT GCCTCCAAACG GGCTGTCGGC GCGAACCGGG CGTGCAGCAGG
 53161 AGCGGGTGTAT CGGCACGCCG CTGGCCAAACG CGGACTCAC CCGCCGCGAC GTGGACGCC
 45 53221 TCGAGGGCCA CGGCACCGGC ACCAGGCTGG GCGACCCCAT CGAGGCACAG GCCGTGCTGG
 53281 CCACCTACGG GCAGGGCGC GACACCCCTG TGCTGCTGGG CTCCGCTGAAG TCCAACATCG
 53341 GCCACACCCA GCCCCCGCGC GGCGTCCCGC GTGTCATCAA GATGGTCCTC GCCATCGCGC
 53401 ACGGCACCCCT GCCCCGCACC CTGCACGTGG ACACGCCGTC CTCCGACGTC GACTGGACGG
 53461 CGGGCGCCGT CGAACCTCCTC ACCGACGCC GGCCTGGCC CGAAACCGAC CGCCCACGGC
 53521 GCGCCGGTGT CTCCCTCTTC GGCGTCAGCG GCACCAACGC CCACATCATC CTCGAAAGCC
 50 53581 ACCCCCCGACC GCCCCCGGAA CCCGCCCCGG CACCCGACAC CGGACCGCTC CGCGTCTGCG
 53641 TCTCGGGCCCG CACCCCGCAG GCACTCGACG CACAGGTACA CGGCCTGCGC GCGTTCTCG
 53701 ACGACAACCC CGGCGCGGAC CGGGTCGCCG TCGCGCAGAC ACTCGCCCGG CGCACCCAGT
 53761 TCGAGCACCG CGCCGTGCTG CTCGGCGACA CGCTCATCAC CGTGAGCCCG AACGCGGCC
 53821 GCGGACCGGT GGTCTTCGTC TACTCGGGC AAAGCACGGT GCACCCGAC ACCGGGGCGC
 55 53881 AACTCGCGTC CACCTACCCC GTGTTCCCG AAGCGTGGCG CGAGGCCCTC GACCACTCG
 53941 ACCCCCCACCA GGGCCCGGCC ACGCACTTCG CCCACCAAGAC CGCGCTCAC GCGCTCTGC
 54001 STCCTGGGG CATCACCCCG CACGCGGTCA TCGGCCACTC CCTCGGTGAG ATCACCGCCG
 54061 CGCACGCCGC CGGTGTCTG TCCCTGAGGG ACGGGGCGC GCTCTCAC ACCCGCACCC
 54121 GCCTGATGGA CCAACTGCCG TCGGGGCGCG CGATGGTCAC CGTCTGACC AGCGAGGAAA
 60 54181 AGGCACGCCA GGTGCTGCCG CGGGCGTGG AGATCGCCGC CGTCAACGGC CCCCCACTCCCC

54241 TCGTGTGTC CGGGGACGAG GAAGCCGTAC TCGAAGCCGC CCGGCAGCTC GGCATCCACC
 54301 ACCGCCTGCC GACCCGCCAC GCCGGCCACT CCGAGCGCAT GCAGCCACTC GTCGCCCCCC
 54361 TCCTCGACGT CGCCCGGACC CTGACGTACC ACCAGCCCCA CACCGCCATC CCCGGCAGCC
 54421 CCACCAACCGC CGAATACTGG GCGCACCAAGG TCCGCGACCA AGTACGTTTC CAGGCGACA
 54481 CGAGCAGTA CCCGGGCGCG ACGTTCTCG AGATCGCCC CAACCAGGAC CTCTCGCCGC
 54541 TCCTCGACGG CGTTGCCGCC CAGACCGGTA CGCCCGACGA GTGCGGGCG CTGCACACCG
 54601 CGCTCGCGCA GCTCCACGTC CGCCGCGTGC CGATCGACTG GACGCTCGTC CTCGGCGGGG
 54661 ACCGGCGGCC CGTCACCGTG CCCACGTATC CGTCCAGCA CAAGGACTAC TGGCTCGGGC
 54721 CCACCTCCCG GGCGATGTG ACCGGCGCGG GGCAGGAGCA GTGCGCAGAC CCGCTGCTCG
 10 54781 GCGCCCGGGT CGCGCTGCC GGCACGGCG GAGTCGTCCT GACCGCCCGC CTGTCGCTGG
 54841 CCTCCCATCC GTGGCTCGGC GAGCACCGG TCGACGGCAC CGTGCCTCTG CCCGGCGCGG
 54901 CCTTCCTCGA ACTCGGGCGC CGCCCGGGG ACGAGGTGCG CTGCGACCTG CTGCACGAAC
 54961 TCGTCATCGA GACGCCGTC GTGCTGCCCG CGACCGGCGG TGTCGCGGTC TCCGTCGAGA
 55021 TCGCCGAACC CGACGACACG GGGCGCGGG CCGTCACCGT CACAGCGGG GCCGACGGCT
 15 55081 CGGGCCTGTG GACCCGACAC GCGGGCGGAT TCCCTGGCAC GGCACCGGCA CCGGCACCGG
 55141 CCACGGACCC GGCACCCCTGG CGCCCGCGG AAGCCGGACC GTGCGACGTC GCGGACGTCT
 55201 ACGACCGGTT CGAGGACATC GGGTACTCCT ACGGACCGGG CTTCCGGGGG CTGCGGGCCG
 55261 CCTGGCGCGC CGGGCACACC GTGTACGCCG AGGTGCGGCT CCCCGACGAG CAGAGCGCCG
 55321 ACGCCGCCCG TTTCACGCTG CACCCCGCGC TGCTCGACGC CGCGTTCCAG GCCGGCGCGC
 20 55381 TGGCCCGCCT CGACGACACC GGGGGGCGG CCGACTGCC GTTCTCGTT CAGGACGTCC
 55441 GCATCCACGC GGCGGGGGCG ACACGGCTGC GGGTCACGGT CGGCCGCGAC GGCAGACGCA
 55501 GCACCGTCCG CATGACCGGC CGGGACGGG AGCTGGTGGC CGTGGTCGGT GCCGTGCTGT
 55561 CGCGCCCGTA CGCGGAAGGC TCCGGTGACG GCCTGCTGCG CCCGGTCTGG ACCGAGCTGC
 55621 CGATGCCCGT CCCGTCCCG GACGATCCGC CGGTGGAGGT CCTCGGCGCC GACCCGGCG
 25 55681 ACGGGACGT TCCGGCGGCC ACCCGGGAGC TGACCGCCCG CGTCCTCGGC GCGCTCCAGC
 55741 GCCACCTGTC CGCCGCGAG GACACCACT TGGTGGTACG GACCGGCACC GGCCCGGCCG
 55801 CTGCCGCCGC CGCGGGCTCG GTCCGCTCGG CGCAGGGGAA GAAACCCGGC CGCGTCGTGC
 55861 TCGTCGAGGC GTCCCCGGAC ACCTCGGTGG AGCTGCTCGC CGCGTGCACG GCGCTGGACG
 55921 AACCGCAGCT GGCGTCCCG GACGGCGTGC TCTTCGCGCC GCGGCTGGTC CGGATGTCCG
 30 55981 ACCCGCGCA CGGCCCCGCTG TCCCTGCCGG ACACCGACTG GCTGCTCACC CGGTCCGCCT
 56041 CGGGCACGTT GCACGACGTC GCGCTCATAG CCGACGACAC GCCCCGGCGG GCGCTCGAAG
 56101 CGGGCGAGGT CGCAGATCGAC GTCCGCGCGG CGGGACTGAA CTTCGCGAT GTGCTGATCG
 56161 CGCTCGGGAC GTACACCGGG GCCACGGCCA TGGCGCGGA GGGCGGGGG GTCGTGGTGG
 35 56221 AGACCGGGCC CGGGCGTGGAC GACCTGTCCC CGGGCGACCG GGTGTTCGGC CTGACCCGGG
 56281 GCGGCATCGG CCCGACGGCC GTCACCGACC GGCCTGGCT GGGCCGGATC CCCGACGGCT
 56341 GGAGCTTCAC CACGGCGCG TCCGTCCCGA TCGTGTTCGC GACCGCGTGG TACGGCCTGG
 56401 TCGACCTCGG CACACTGCGC GCCGGCGAGA AGGTCTCGT CCACGCGGCC ACCGGCGGTG
 56461 TCGGCATGGC CGCGCACAG ATCGCCCGCC ACCTGGCGC CGAGCTCTAC GCCACCGCCA
 40 56521 GTACCGGCAA CGACGACGTC CTGCGCGCCG CGGGCTGCC CGACACGAC ATCGCCGACT
 56581 CTCGGACGAC CGCGTTCGGG ACCGCTTCC CGCGCATGGA CGTCGTCTG AACCGCCTGA
 56641 CGGGCGAGTT CATGACCGCG TCGCTCGACC TGCTGGACG CGACGGCCGG TTCGTCGAGA
 56701 TGGGCCGCAC CGAGCTGCGC GACCCGGCG CGATCGTCCC CGCCTACCTG CGGTCGACC
 56761 TGCTGGACGC GGGCGCCGAC CGCATCGCG AGATCTGGG CGAACCTGCTC CGGCTGTTCG
 56821 ACGCGGGCCG GCTGGAGGCCG CTGCCGGTCC GTGCCCTGGG CGTCCGGCAG SCACCGACG
 45 56881 CGCTCGGCTG GATGAGCCCG GCCCCCACA TCGGCAAGAA CGTCTGACG CTGCCCCGGC
 56941 CGCTCGACCC GGAGGGCGCC GTCGTCTCA CGGGCGCTC CGGCACGCTC GCCGGCATCC
 57001 TCGCCCGCA CCTGCGCGAA CGGCATGTCT ACCTGCTGTC CGGGACGGCA CCGCCCGAGG
 57061 GGACGGCCGG CGTCCACCTG CCCTGCACG TCGGTGACCG GGACCGCTG CGGGCGGGCC
 57121 TGGAGCGGGT GGACCGGGCG ATCACCGCCG TGGTGCACCT CGCCGGTGCG CTGGACGACG
 50 57181 GCACCGTCGC GTCGCTCACC CCCGAGCGTT TCGACACGGT GCTGCGCCCG AAGGCGACG
 57241 GCGCTGGTA CCTGCACCGAG CTGACGAAGG AGCAGGACCT CGCCGCGTTC GTGCTCTACT
 57301 CGTCGGCGC CGGGCGTGCCTG GGCAACGCCG GCCAGGGCAA CTACGTCGCC GCGAACCGCT
 57361 TCCTCGACGC GCTCGCCGAG CTGCACCGACG GTTCCGGCT GCCGGCCCTC TCCATCGCT
 57421 GGGGCTCTG GGAGGACGTG AGCGGGCTCA CGCGCGCGT CGGCAGAGCC GACCGGGACC
 55 57481 GGATGCGCG CAGCGGTTTC CGGGCCATCA CGCGCAACA GGGCATGCAC CTGTACGAGG
 57541 CGGCGGGCCG CACCGGAAGT CCCGTGGTGG TCGCGGGCGC GCTCGACGAC GCGCCGGACG
 57601 TGCCGCTGCT GCGCGGCCG CGGGCGACGA CGTCCGGCG GGGCGCCGTC CGGGAGTGGT
 57661 CGTCCGCCGA CGGGCTCGCC GCGCTGACCG GCGACGAGCT CGCCGAAGCG CTGCTGACGC
 57721 TCGTCGGGA GAGCACCGCC GCGCTGCTCG GCCACGTGGG TGGCGAGGAC ATCCCCCGCA
 60 57781 CGGGGGCGTT CAAGGACCTC GGCATCGACT CGCTCACCGC GGTCCAGCTG CGAACCGCCC

	57841	TCACCGAGGC	GACCGGTGTG	CGGCTGAACG	CCACGGCGGT	CTTCGACTTC	CCGACCCCCG
	57901	ACGTGCTCGC	CGGGAAAGCTC	GGCGACGAAC	TGACCGGCAC	CCCGCGGCC	GTCGTGCCCC
	57961	GGACCCGCGC	CACGGCCGGT	GCGCACGACG	AGCCGCTGGC	GATCGTGGGA	ATGGGCTGCC
	58021	GGCTGCCCGG	CGGGGTGCGG	TCACCCGAGG	AGCTGTGGCA	CCTCGTGGCA	TCCGGCACCG
5	58081	ACGCCATCAC	GGAGTTCCCG	ACGGACCGCG	GCTGGGACGT	CGACCGCAGTC	TACGACCCGG
	58141	ACCCCGACGC	GATCGGCAAG	ACCTTCGTCC	GGCACGGTGG	CTTCCTCACC	GGCGCGACAG
	58201	GCTTCGACGC	GGCGTTCTTC	GGCATCAGCC	CGCGCGAGGC	CCTCGCGATG	GACCCGCAGC
	58261	AGCGGGTGCT	CCTGGAGACG	TCGTGGGAGG	CGTTGAAAG	CGCCGGCAGTC	ACCCCGGACT
	58321	CGACCCGCGG	CAGCGACACC	GGCGTGTTCG	TCGGCGCCTT	CTCCTACGGT	TACGGCACCG
10	58381	GTGCGGACAC	CGACGGCTTC	GGCGCGACCG	GCTCGCAGAC	CAGTGTGCTC	TCCGGCCGGC
	58441	TGTCTGACTT	CTACGGTCTG	GAGGGTCCGG	CGGTACCGGT	CGACACGGCG	TGTTCTCGT
	58501	CGCTGGTGGC	GCTGCACCAAG	GCCGGGCAGT	CGCTCGCTC	CGGCGAATGC	TCGCTCGCCC
	58561	TGGTCGGCGG	CGTCACGGTG	ATGGCGTCTC	CCGGCGCTT	CGTGGAGTT	TCCCGGCAGC
15	58621	GCGGCCTCGC	GCCGGACGGC	CGGGCGAAGG	CGTTCGGC	GGGTGCGGAC	GGCACGAGCT
	58681	TCGCCGAGGG	TGCCGGTGTG	CTGATCGTCG	AGAGGCTCTC	CGACGCCGAA	CGCAACGGTC
	58741	ACACCGTCC	GGCGGTGTC	CGTGGTTCGG	CGGTCAACCA	GGATGGTGCC	TCCAACGGGC
	58801	TGTGGCGGCC	GAACGGGCCG	TCGCAGGAGC	GGGTGATCCG	GCAGGCCCTG	GCCAACGCCG
	58861	GGCTCACCCCC	GGCGGACGTG	GACGCCGTCG	AGGCCCCACGG	CACCGGCACC	AGGCTGGGCG
20	58921	ACCCCATCGA	GGCACAGGGC	GTACTGGCA	CCTACGGACA	GGAGCGCGCC	ACCCCCCTGC
	58981	TGCTGGGCTC	GCTGAAGTCC	AACATCGCC	ACGCCCAGGC	CGCGTCCGGC	GTCGCCGGCA
	59041	TCATCAAGAT	GGTGCAGGCC	CTCCGGCACG	GGGAGCTGCC	GGCGACGCTG	CACGCCGACG
	59101	AGCCGTGCGC	GCACGTGAC	TGGACGGCGC	GGCCCGTCGA	ACTGCTGACG	TCGGCCCGGC
	59161	CGTGGCCCGA	GACCGACCGG	CCACGGCGTG	CCGCCGTCTC	CTCGTTCGGG	GTGAGCGGCA
25	59221	CCAACGCCA	CGTCATCTG	GAGGCCGGAC	CGGTAACGGA	GACGCCGGC	GCATCGCCTT
	59281	CGGGTGACCT	TCCCCCTGCTG	GTGTGGGAC	GCTCACCGGA	AGCGCTCGAC	GAGCAGATCC
	59341	GCGCAGTCGCG	CGCCTACCTG	GACACCAACCC	CGGACGTGGA	CGGGGTGGCC	GTGGCACAGA
	59401	CGCTGGCCCG	GCGCACACAC	TTCCCCCACC	CGGCCGTGCT	GCTCGGTGAC	ACCGTCATCA
	59461	CCACACCCCC	CGCGGACCCG	CCCACGAAC	TCGTCTCGT	CTACTCCGGC	CAGGGCACCC
30	59521	AGCATCCCAC	GATGGGGCAG	CAGCTCGCCG	CCGCCCATCC	CGTGTTCGCC	GACGCCCTGGC
	59581	ATGAAGCGCT	CCGCCGCC	GACAACCCCC	ACCCCCACGA	CCCCACGCA	AGCCAGCATG
	59641	TGCTCTTCGC	CCACCAGGGC	GGCTTCACCG	CCCTCCTGCG	GTCTCTGGGGC	ATCACCCCGC
	59701	ACGCGGTCA	CGGCCACTCG	CTGGGCGAGA	TCACCGCGC	GCACGCCGCC	GGCATCCTGT
	59761	CGCTGGACGA	CGCGTGCACC	CTGATCACCA	CGCGCGCCCG	CCTCATGCAC	ACGCTCCCAC
35	59821	CACCCGGTGC	CATGGTCACC	GTACTGACCA	CGAAGAGAAA	GGCACGCCAG	GCGTTGCCG
	59881	CGGGCGTGG	GATGCCGCC	GTCAACGGGC	CCCACTCCAT	CGTGTGTGCC	GGGGACGAGG
	59941	ACGCGGTGCT	CACCGTCGCC	GGGCAGCTCG	GCATCCACCA	CCGCCTGCC	GCCCCGCACG
	60001	CCGGGCACTC	CGCGCACATG	GAGCCCGTGG	CCGCCGAGCT	GCTCGCCACC	ACCCGCCGGC
	60061	TCCGCTACCA	CCCTCCCCAC	ACCTCCATT	CGAACGACCC	CACCACCGCT	GAGTACTGGG
40	60121	CCGAGCAGGT	CCGCAAGGCC	GTGCTGTCC	ACGCCACG	GCAGCAGTAC	CCGGACGCCG
	60181	TGTTCTGTG	GATCGGCC	GCCCAGGACC	TCTCCCCGCT	CGTCGACGGG	ATCCCCGTC
	60241	AGAACGGCAC	CGCGGACGAG	GTGACCGC	TGCACACC	GCTCGCGCAC	CTCTACCGC
	60301	GCGGTGCCAC	GCTCGACTGG	CCCCGCATCC	TCGGGGCTGG	GTACGGCAC	GACGCCGGATG
	60361	TGCCCCGCTA	CGCGTTCAA	CGCGGCCACT	ACTGGATCGA	GTGCGCACGCC	CCGGCCGCAT
45	60421	CCGACCGGGG	CCACCCCGTG	CTGGGCTCCG	GTATCGCCCT	CGCCGGGTG	CCGGGCCGGG
	60481	TGTTCACGGG	TTCCGTGCCG	ACCGGTGCGG	ACCGCGCGG	GTTCGTCGCC	GAGCTGGCGC
	60541	TGGCCGCC	GGACCGGGTC	GAATCGGCCA	CGGTCGAGCG	GCTCGACATC	GCCTCCGTG
	60601	CCGGCCGGCC	GGGCCATGGC	CGGACGA	TACAGACCTG	GGTCGACGAG	CCGGCCGGACG
	60661	ACGGCCGGCG	CGGGTCA	GTGACACCC	GCACCGCGA	CGCCCCGTG	ACGCTGCACG
	60721	CCGAGGGGT	GCTGCGCCCC	CATGGCACGG	CCCTGCCGA	TGCGGCCGAC	GCCGAGTGGC
50	60781	CCCCACCGGG	CGCGGTGCC	CGGGACGGGC	TGCCGGGTG	GTGGCGCCGG	GGGGACCAAGG
	60841	TCTTCGCCGA	GGCCGAGGTG	GACGGACCGG	ACGGTTTCGT	GGTGCACCCCC	GACCTGCTCG
	60901	ACGCGGTCTT	CTCCGCGGTC	GGCGACGGAA	GCCGCCAGCC	GGCCGGATGG	CGCGACCTGA
	60961	CGGTGCACGC	GTCGGACGCC	ACCGTACTGC	GCGCCTGCC	CAACCGCGC	ACCGACGGAG
55	61021	CCATGGGATT	CGCCGCC	GACGGCGCC	GCCTGCCG	ACTCACCGC	GAGGCGGTGA
	61081	CGCTGCCGGA	GGTGGCGTC	CCGTCGGC	CCGAGGAGTC	GGACGGCCTG	CACCGTTGG
	61141	AGTGGCTCGC	GGTCGCCGAG	CGGGTCTACG	ACGGTACCT	GGCCGAGGGA	CATGTCCTGA
	61201	TCACCGCCG	CCACCCGAC	GACCCCGAGG	ACATACCCAC	CCGCGCC	ACCCGCCA
	61261	CCCGCGTCT	GACCGCCCTG	CAACACCAC	TCACCAC	CGACCCACACC	CTCATCGTCC
	61321	ACACCAC	CGACCCCGCC	GGCGCCACCG	TCACCGGC	CACCCGCACC	GCCCAGAACG
60	61381	AACACCCCCA	CCGCATCCG	CTCATCGAA	CCGACCA	CCACACCCCC	CTCCCCCTGG

61441	CCCAACTCGC	CACCCCTCGAC	CACCCCCCACC	TCCGCCTCAC	CCACCACACC	CTCCACCACC
61501	CCCACCTCAC	CCCCCTCCAC	ACCACCACCC	CACCCACCCAC	CACCCCCCTC	AACCCCGAAC
61561	ACGCCATCAT	CATCACCGGC	GGCTCCGGCA	CCCTCGCCGG	CATCCTCGCC	CGCCACCTGA
61621	ACCACCCCCA	CACCTACCTC	CTCTCCCGCA	CCCCACCCCC	CGACGCCACC	CCC GG CACCC
5	61681	ACCTCCCCCTG	CGACGTGGC	GACCCCCACC	AACTCGCCAC	CACCCCTCACC
61741	AACCCCTCAC	CGCCATCTTC	CACACC GCCG	CCACCCCTCGA	CGACGGCATE	CTCCACGCC
61801	TCACCCCCGA	CCGCCTCACC	ACCGTCTTCC	ACCCCAAAGC	CAACGCCGC	TGGCACCTGC
61861	ACCACCTCAC	CCAAAACCAA	CCCCTCACCC	ACTTCGTCTT	CTACTCCAGC	GCCGCCGCCG
61921	TCCTCGGCAG	CCCCGGACAA	GGAAACTACG	CCGCCGCCAA	CGCCCTCCTC	GACGCCCTCG
10	61981	CCACCCACCG	CCACACCCCTC	GGCCAACCCG	CCACCTCCAT	CGCCCTGGGGC
62041	CCACCA CGCAC	CCTCACCGGA	CAACTCGACG	ACGCCGACCG	GGACCGCATIC	CGCCGGCGCG
62101	TTTCTCTCCC	GATCACGGAC	GACGGAGGGCA	TGCGCCCTCA	CGAGGGGGCC	GTCGGCTCCG
62161	GCGAGGACTT	CGTCATGGCC	GCCCGATGG	ACCCGGCACA	GCGATGACC	GGCTCCGTAC
15	62221	CGCCCCATCCT	GAGCGGCCCTG	CGCAGGAGCG	CGCGGCCGCGT	CGCCCGTGCC
62281	TCGCCCCAGCG	GCTCGCCGAG	CTGGCCGACG	CCGACCCGCG	CGCGGCCGCTG	ACCACCCCTCG
62341	TCTCGGACGC	CACGGCCGCC	GTGCTCGGCC	ACGCCGACGC	CTCCGAGATC	GCGCCGACCA
62401	CGACGTTCAA	GGACCTCGGC	ATCGACTCGC	TCACCGCGAT	CGAGCTGC	AACCGGCTCG
62461	CGGAGGCGAC	CGGGCTGC	CTGAGTGC	CGCTGGTGT	CGACCACCCG	ACACCTCGGG
20	62521	TCCTCGCCGC	CAAGCTCCGC	ACCGATCTGT	TCGGCACGGC	CGTGGCCACG
62581	CGGCACGGAC	CCACCA CGAC	GAGCCACTCG	CGATCGTGG	CATGGCGTGC	CGACTIGCCCG
62641	GCGGGGTCGC	CTCGCCGGAG	GACCTGTGGC	AGCTCGTGG	GTCCGGCACC	GACGCGATCA
62701	CCGAGTTCCC	CACCGACCGC	GGCTGGGACA	TCGACCGGCT	GTTCGACCCG	GACCCGGACG
62761	CCCCCGGCAA	GACCTACGTC	CGGCACGGCG	GCTTCCTCGC	CGAGGGCGCC	GGCTTCGATG
25	62821	CCGCGTTCTT	CGGCATCAGC	CGCGCGGAGG	CACGGGCCAT	GGACCCGAG
62881	TCCTCGAAC	CTCCTGGGAG	GCGTTCGAGA	ACGCGGGCAT	CGTGGCCGAC	ACGCTCGCG
62941	GCAGCGACAC	CGGCGTGTTC	ATGGGCGCGT	TCTCCCATGG	GTACGGCGCC	GGCGTCGACC
63001	TGGGGGGGTT	CGGCGCCACC	GCCACGCAGA	ACAGCGTGCT	CTCCGGCCGG	TTGTCTGTACT
63061	TCTTCGGCAT	GGAGGGCCCG	GCCGTCACCG	TCGACACCGC	CTGCTCGTGC	TCGCTGGTGC
30	63121	CCCTGCACCA	GGCGGCACAG	GCGCTCGGGA	CTGGAGAATG	CTCGCTGGCG
63181	GTGTACGGT	GATGCCACC	CCGCTGGGCT	ACGTCGAGTT	CTGCGGCCAG	CGGGGACTCG
63241	CCCCCGACGG	CCGTTGCCAG	GCCTTCGCG	AAGGCGCCGA	CGGCACGAGC	TTCTCGGAGG
63301	GCGCCGGCGT	TCTTGTGCTG	GAGCGGCTCT	CCGACGCCGA	GCGAACCGGA	CACACCGTCC
35	63361	TCGCGGTGCGT	CCGCTCCTCC	GCCGTCAACC	AGGACGGCGC	CTCCAACGGC
63421	CCAACGGCCC	CTCCCAGCAG	CGCGTCATCC	GCCAGGGCCT	CGACAAGGCC	GGGCTCGCCC
63481	CCGCCGACGT	GGACGTGGTG	GAGGCCACG	GCACCGGAAC	CCC GCTGGGC	GACCCGATCG
63541	AGGCACAGGC	CATCATCGCG	ACCTACGGCC	AGGACCGCGA	CACACCGCTC	TACCTCGGTT
40	63601	CGGTCAAGTC	GAACATCGGA	CACACCCAGA	CCACCGCCGG	TGTGCCGGC
63661	TGGTCATGGC	GATGCGCCAC	GGCATCGCG	CGAAGACACT	GCACGTGGAC	GAGCCGTCGT
63721	CGCATGTGGA	CTGGACCGAG	GGTGCCTGG	AACTGCTCAC	CGAGGCAGG	CCGTGGCCCG
45	63781	ACGCGGGACG	CCC CGCGCCG	GCGGGCGTGT	CGTCGCTCGG	TATCAGCGGT
63841	ACGTGATCCT	TGAGGGTGT	CCC GGGCCGT	CGCGTGTGG	GCCGTCTGT	GACGGGTTGG
63901	TGCCGTTGCC	GGTGTGCGT	CGGAGTGAGG	CGAGTCTGCG	GGGGCAGGTG	GAGCGGCTGG
63961	AGGGGTATCT	GCGCGGGAGT	GTGGATGTGG	CCGCGGTGCG	CAAGGGTTG	GTGCGTGAGC
64021	GTGCTGTCTT	CGGTACCGT	GCGGTACTGC	TGGGTGATGC	CCGGGTGATG	GGTGTGGCGG
50	64081	TGGATCAGCC	GGGTACGGTG	TTCTGCTTT	CCGGGCAAGGG	TGCTCAGTGG
64141	GTGTGGAGTT	GATGGACCGT	TCTCGCGTGT	TCGCGGCTCG	TATGGAGGAG	TGTGCGCGGG
64201	CGTTGTTGCC	GCACACGGGC	TGGGATGTGC	GGGAGATGT	GGCGCGGCCG	GATGTGGCGG
64261	AGCGGGTGG	GGTGGTCCAG	CGGGCCAGCT	GGGCGGTGCG	GGTAGCCTG	GCCGCACTGT
64321	GGCAGGCCCA	CGGGGTGCGA	CCCGACGCCG	TGATCGAC	CTCCCAGGGC	GAGATCGCGG
55	64381	CGGCGTGC	GGCGGGGGCC	CTCAGCCTTG	AGGACGCCGC	CCCGGTGGTG
64441	GCCAGGTAT	CGCGGCCGGA	CTGGCCGGGC	GGGGAGCGAT	GGCTTCGGTG	GCATTGCCG
64501	CGGGTGTAGG	CGGTCTGGC	GAGGGCGTGT	GGATCGCGGC	GCGTAACGGC	CCC GCTCGA
64561	CAGTCGTGGC	CGCGGAGGCC	TCGGCGGTGG	AGGACGTGGT	GACCGGGTAT	GAGACCGAAG
64621	GCGTGCAGT	GGTCGTATC	GGCGTCACT	ACGCCCTCCA	CACGCCAAC	GTGGAAGCCA
55	64681	TCGAGGACGA	ACTCGCTGAG	GTACTGAAGG	GAGTTGCAGG	GAAGGCCGCG
64741	GGTGGTCGAC	CGTGGACAGC	GCCTGGGTGA	CCGAGCCGGT	GGATGAGAGT	TACTGTG
64801	GGAACCTGCG	TGGCCCCGTC	CGCCTGGACG	CGGCGGTGGC	GGAGCTGGAC	GGGTCCGTG
64861	TGCTGGAGTG	CAGCGCCCAT	CCGGTGTGC	TGCCGGCGAT	GGAACAGGCC	CACACGGTGG
64921	CGTCGTTGCC	CACCGGTGAC	GGCGGCTGGG	AGCGATGGCT	GACGGCGTTG	GCGCAGGCCGT
60	64981	GGACCCCTGGG	CGCGGCAGTG	GACTGGGACA	CGGTGGTCGA	ACCGGTGCCA
						GGGC GGCTGC

65041 TCGATCTGCC CACCTACGCG TTGAGCGCC GGCCTACTG GCTGGAAGCG GCCGGTGCCA
 65101 CCGACCTGTC CGCGGCCGGG CTGACAGGG CAGCACATCC CATGCTGGCC GCCATCACGG
 65161 CACTACCCGC CGACGACGGT GGTGTTGTC TCACCGGCCG GATCTCGTT CGCACGCATC
 65221 CCTGGCTGGC TGATCACGCG GTGCGGGCA CGGTCTGCT GCCGGGCACG GCCTTGTGG
 5 65281 AGCTGGTCAT CGGGGCCGGT GACGAGACCG GTTGCAGGAT AGTGGATGAA CTGGTCATCG
 65341 AATCCCCCT CGTGGTGCAG GCGACCGCAG CGTGGATCT GTCGGTGACC GTGGAAGGAG
 65401 CTGACGGAGC CGGACGGCG CGAGTGACCG TCCACGCCG CACCGAAGGC ACCGGCAGCT
 65461 GGACCCGGCA CGCCAGCGGC ACCCTGACCC CGACACCCCC CGACACCCCC AACGCTTCCG
 65521 GTGTTGTCGG TGGGGAGCCG TTCTCGACT GGCACCTGCACTGCCGCG GCCGTCGACA
 10 65581 CCTCGGAGTT CTACTTGCCT CGGACCGCAG TGGCTACCG GTTCGGACCC ATGTTCCGCG
 65641 GAATGCGGGC TGCCTGGCGT GATGGTGACA CGTGTACGC CGAGGTCGCG CTCCCCGAGG
 65701 ACCGTGCCGC CGACGCGGAC GTTTCGGCA TGACCCGGC GTCACCGCAG GCACCGCTTGC
 65761 AGAGCGGCAG CCTGCTCATG CTGGAATCGG ACGGGAGCA GAGCGTCAA CTGCGCTTCT
 65821 CCTGGCACGG CGTCCGGTTC CACCGACGG GCGGACCAT GTCGCGGTG GCGGTCGTC
 15 65881 CGGGCCCGGA CGGCCTCCGG CTGCACTGCCG CGGACAGCGG GAACCGTCCC GTCGGACGAG
 65941 TCGACCGCCT CGTGAACCCGG TCCCCGGAAAG CGGACCTGCG GCCCCGGAT CCGATGCTGC
 66001 GGGTGGGTG GGCCCCGGTGC CGGCTACCTG CGGGGGCCGG TCCGTCGAC GCGGACGTGC
 66061 TGACGCTGCG CGGCGACGAC GCCGACCCGC CGGGGGAGAC CGGGGACCTG ACCACCCGTG
 66121 TTCTCGACGC GTCGCTCCGG GCCGACCGGC CGGTGATCTT CCAGGTGACCG GGTGGCTCG
 20 66181 CCGCCAAGGC GGCGCAGGG CTGGTCCGCA CGGCTCAGAA CGAGCAGCCC GGCGCTTCT
 66241 TCCTCGTCGA AACGGACCCG GGAGAGGTCC TGGACGGCGC GAAGCGGAC GCGATCGCG
 66301 CACTCGCGA GCCCCATGTG CGGCTCGCG ACGGCCTT CGAGGCAGCC CGGCTGATGC
 66361 GGGCCACGCC GTCCCTGACG CTCCCGGACA CGGGGCTGTG GCAGCTGCGG CGTCCGCCA
 66421 CCGGTTCCCT CGACGACCTT CGCGTGTCC CGACGACGC CCCGGACCGG CGCGTCCGG
 25 66481 CGGGCGAGGT CGGGATCGCG GTACCGCGG CGGGCCTGAA CTTCGGGAT GTCACGGTCC
 66541 CGCTGGTGT GTCGCCGAT GCGCGTCCGC TCGCGACGA GGCGCGGGT GTCGCTCTGG
 66601 AGACCGGCC CGGTGTGCAAC GACCTGGCGC CGGGCGACCG GGTCTGGGG ATGCTCGGG
 66661 GCGCCTCGG ACCGGTGCACG ATCACCGAAC GGCGGCTGT CGGGCGGATG CGGACGGCT
 66721 GGACGTTCCC GCAGGGGGCG TCCGTATGA CGCGTTCGC GACCGCGTGG TACGGCTGG
 30 66781 TCGACCTGGC CGGGCTGCC CGCGCGAGA AGGTCTGTAT CCACCGGGCG GCGACGGTG
 66841 TCGGCGCGGC GGCGTCCAG ATCGCGCGC ATCTGGGCGC GGAGGTGTAC GCGACCAACCA
 66901 GCGCCGCGAA GCGCCATCTG GTGGACCTGG ACGGAGCGCA TCTGGCGAT TCCCGCAGCA
 66961 CCGCGTTCGC CGACGCGTTC CGGCCGGTGC ATGTCGTGT CAACTCGCTC ACCGGTGAAT
 67021 TCCTCGACGC GTCCGTGGC CTGCTCGCGG CGGGTGGCG GTCATCGAG ATGGGAAAGA
 35 67081 CGGACATCCG GCACGCCGTC CAGCAGCCGT TCGACCTGTAT GGACGCCGGC CCCGACCGGA
 67141 TGCAGCGGAT CATCGTCGAG CTGCTCGGCC TGTTCGCGCG CGACGTGCTG CACCCGCTGC
 67201 CGGTCCACGC CTGGGACGTG CGGCAGCGC GGGAGGGCGTT CGGCTGGATG AGCAGCGGGC
 67261 GTCACACCGG CAAGCTGGTG CTGACGGTCC CGCGGCCGCT GGATCCCGAG GGGCGCTCG
 67321 TCATCACCGG CGGCTCCGG ACCCTCGCCG GCATCTCGC CGGCCACCTG GGCCACCCCC
 40 67381 ACACCTACCT GCTCTCCCG ACCCCACCCC CGACACCCAC CCCGGCACC CACCTCCCT
 67441 GCGACGTCGG CGACCCCCAC CAACTCGCCA CCACCTCGC CGCATTCCCC CAACCCCTCA
 67501 CCGCGTCTT CCACACCGCC GGAACCTCG ACACGCCCT GTCGACAAC CTCACCCCC
 67561 ACCCGCTCGA CACCGCTCTC AAACCCAAGG CCGACGCCGC CTGGCACCTG CACCGGCTCA
 67621 CCCCGACAC CGACCTCGCC GCGTCGTG TCTACTCCGC GGTCGCCGGC CTCATGGCA
 45 67681 GCGGGGGCA GGGCAACTAC GTCGCGCGA ACGCGTTCT CGACCGCCTC CGCGAACACC
 67741 GCGGTGCGCA AGGGCTGCC GCGCAGTCCC TCGCATGGGG CATGTGGCG GACGTGAGCG
 67801 CGCTCACCGC GAAAATCACCG GACCGGGACC GCCAGCGCAT CGGCGCAGC GGATTCCCGC
 67861 CGTTGAGCGC CGCGGACGGC ATGCGGCTGT TCGACGCGC GACCGTACCG CGGGAACCGG
 67921 TCGTCGTGCG GACGACCGTC GACCTCACCC AGCTCGACGG CGCCGTGCGC CGGTTGCTCC
 50 67981 CGGGCTGGC CGCGCACCGG CGCGGCCGG CGCGCACCGT CGCCCGCAAC GCGGCGAAG
 68041 AGCCCCCTGGC CGTGCCTCTT GCGGGCGTA CGCGGCCGA GCACGCCGC ATCATGCGG
 68101 AGGTGCTGCT CCGCCACCGC CGCGGGTCC TCGCGTACCG GCTGGGCGAC CGCGTGGCGG
 68161 CGGACCGTCC GTTCCCGAG CTCGGTTTCG ATTGCGTGTAC CGCGGTGCGAC CTGCGCAATC
 68221 GGCTCGCGC CGAGACGGGG CTGCGCTGC CGACGACGCT GGTGTTGAGC CACCCGACGG
 55 68281 CGGAGGGCGT CACCGCCACCG CTGCTCGACC TGATCGACG TCCACCGCC CGGATCGCCG
 68341 GGGACTCCCT GCGCGCGGTG ACGGCCGCTC CGTGGCGGC CGCGCGGGAC CAGGACGAGC
 68401 CGATCGCCAT CGTGGCGATG CGTGGCGGC TGCCCGGTGG TGTGACGTCG CCCGAGGACC
 68461 TGTGGCGGCT CGTCGAGTCC GGCACCGACG CGATCACCAAC GCCTCTGAC GACCGCGGCT
 68521 GGGACGTCGA CGCGCTGTAC GACGCGGACCG CGGACGCCGC CGGCAAGGCG TACAACCTGC
 60 68581 GGGCGGGTTA CCTGGCCGGG CGGGCGGAGT TCGACGCGGC GTTCTCGAC ATCAGTCCGC

	68641	GCGAAGCGCT	CGGCATGGAC	CCGCAGCAAC	GCCTGCTGCT	CGAAACGGCG	TGGGAGGCGA
	68701	TCGAGCGCG	CCGGATCAGT	CCGGCGTCGC	TCCGCGCCG	GGAGGTCGGC	GTCTATGTCG
	68761	GTGCGCCCG	GCAGGGCTAC	GGGCTGGCG	CCGAGGACAC	CGAGGGCCAC	GCGATCACCG
5	68821	GTGGTCCAC	GAGCCTGCTG	TCCGGACGGC	TGGCGTACGT	GCTCGGGCTG	GAGGGCCCCG
	68881	CGGTCAACCGT	GGACACGGCG	TGCTCGTCGT	CTCTGGTCGC	GCTGCATCTG	GCCTGCCAGG
	68941	GGCTGCGCCT	GGGCGAGTGC	GAACTCGCTC	TGGCCGGAGG	GGTCTCCGTA	CTGAGTTCGC
	69001	CGGCCGCGTT	CGTGGAGTTC	TCCCAGCCAGC	GCGGGCTCGC	GGCCGACGGG	CGCTGCAAGT
10	69061	CGTTCGGCGC	GGGCGCGGAC	GGCACGACGT	GGTCCGAGGG	CCTGGGGCTG	CTCGTACTGG
	69121	AACGGCTCTC	CGACGCCGAG	CGGCTCGGGC	ACACCGTGCT	CGCCGTCGTC	CGCGGAGCG
	69181	CCGTACCGTC	CGACGGCGCC	TCCAACGGCC	TCACCGCGCC	GAACGGGCTC	TCGCAGCAGC
	69241	GGGTCATCCG	GAAGGCGCTC	GCCGCGGCCG	GGCTGACCGG	CGCCGACGTG	GACGTCGTG
	69301	AGGGGCACGG	CACCGGCACC	CGGCTCGGGC	ACCCGGTCGA	GGCGGACGCG	CTGCTCGCG
	69361	CGTACGGGCA	GGACCGTCCG	GCACCGGTCT	GGCTGGGCTC	GCTGAAGTCG	AACATCGGAC
15	69421	ATGCCACGGC	CGCGGCCGGT	GTGCGGGCG	TCATCAAGAT	GGTGCAGGCG	ATCGGCGCG
	69481	GCACGATGCC	GGGGACGCTG	CATGTGGAGG	AGCCCTCGCC	CGCCGTCGAC	TGGAGCACCG
	69541	GACAGGTGTC	CCTGCTCGGC	TCCAACCGGC	CCTGGCCGGA	CGACGAGCGT	CGCGGCCGG
	69601	CGGCCGCTCTC	CGCGTTCGGG	CTCAGCGGGA	CGAACCGC	CGTCATCCTG	GAACAGCACC
	69661	GTCCGGCGCC	CGTGGCGTCC	CAGCCGCCCC	GGCCGCCCG	TGAGGAGTCC	CAGCCGCTGC
	69721	CGTGGGTGCT	CTCCCGCGGG	ACTCCGGCCG	CGCTGCGGGC	CCAGGCGGCC	CGGCTGCGCG
20	69781	ACCACCTCGC	GGCGGCACCG	GACCGGGATC	CGTTGGACAT	CGGGTACGCG	CTGGCCACCA
	69841	GCCGCGCCCA	GTTCGCCCCAC	CGTGGCGCG	TCGTCGCCAC	CACCCCGGAC	GGATTCCGTG
	69901	CCGCGCTCGA	CGGCCTCGCG	GACGGCGCG	AGGCGCCCG	AGTCGTCACC	GGGACCGCTC
	69961	AGGAGCGCG	CGTCGCTTC	CTCTTCGACG	GCCAGGGCGC	CCAGCGCGCC	GGAATGGGGC
	70021	GCGAGCTCCA	CCGCCGGTTC	CCCCTCTCG	CCGCCGCGTG	GGACGAGGTC	TCCGACCGT
25	70081	TCGGCAAGCA	CCTCAAGCAC	TCCCCCACGG	ACGTCTACCA	CGGCGAACAC	GGCGCTCTCG
	70141	CCCATGACAC	CCTGTACGCC	CAGGCCGGCC	TGTTCACGCT	CGAAGTGGCG	CTGCTGCGC
	70201	TGCTGGAGCA	CTGGGGGGTG	CGGCCGGACG	TGCTCGTCGG	GAACACTCCGTC	GGCGAGGTGA
	70261	CCGCGGCGTA	CGCGCGCGGG	GTGCTCACCC	TGGCGGACGC	GACGGAGTTG	ATCGTGGCCC
	70321	GGGGCGGGC	GCTGCGGGCG	CTGCCGCCCG	GGGCGATGCT	CGCCGTCGAC	GGAAGCCCGG
30	70381	CGGAGGTGCG	CGCCCGCACG	GATCTGGACA	TCGCCCGGGT	CAACGGCCCG	TCCGCCGTGG
	70441	TGCTGCCCGG	TTCGCCGGAC	GATGTGGCG	CGTTCGAACG	GGAGTGGTCG	GC GGCCGGG
	70501	GGCGCACGAA	ACGGCTCGAC	GTCGGGACCG	CGTCCCACTC	CCGGCACGTC	GACGGTGC
	70561	TGACGGCTT	CCGTACGGTG	CTGGAGTCGC	TCGCGTTCGG	CGCGGCGCG	CTGCCGGTGG
	70621	TGTCACGAC	GACGGGCCGG	GACGCCGCG	ACGACCTCAT	AACGCCCGCG	CACTGGCTGC
35	70681	GCCATGCGCG	TCGGCCGGTG	CTGTTCTCGG	ATGCCGTCGG	GGAGCTGGCG	GACCGCGCG
	70741	TCACCACTGT	CGTGGCGTC	GGCCCCCTCGG	GCTCCCTGGC	GTCCGCCGCG	CGGGAGAGCG
	70801	CGGGGGAGGA	CGCCGGGAC	TACCGCGGG	TGCTGCGCG	CCGGACCGGT	GAGGAGACCG
	70861	CGGCGCTGAC	CGCCCTCGCC	GAGCTGCA	CCACGGCGT	CCCGGTCGAC	CTGGCCGCG
	70921	TACTGGCCGG	TGGCCGGCCA	GTGGACCTTC	CCGTGTACG	GTTCCAGCAC	CGTTCCCTACT
40	70981	GGCTGGCCCC	GGCCGTGGCG	GGGGCGCCGG	CCACCGTGGC	GGACACCGGG	GGTCCGGCGG
	71041	AGTCCGAGCC	GGAGGACCTC	ACCGTCGCCG	AGATCGTCCG	TCGGCCAC	GGGGCCCTGC
	71101	TCGGCGTCAC	GGACCCCGCC	GACGTGATG	CGGAAGCGAC	GTTCTCGCG	CTCGGTTTCG
	71161	ACTCACTGGC	GGTGCAGCGG	CTGCCAAC	AGCTCGCTC	GGCAACCGGG	CTGGACCTGC
	71221	CGGCGGCCGT	CCTGTTCGAC	CACGACACCC	CGGCCGCGCT	CACCGCGTTC	CTCCAGGACC
45	71281	GGATCGAGGC	CGGCCAGGAC	CGGATCGAGG	CGGCCGAGGA	CGACGACGCG	CCCACCGTGC
	71341	TCTCGCTCCT	GGAGGAGATG	GAGTCGCTG	ACGCCGCGGA	CATCGCGCG	ACGCCGCGCC
	71401	CGGAGCGTGC	GGCCATCGCC	GATCTGCTG	ACAAGCTCGC	CCATACCTGG	AAGGACTACC
	71461	GATGAGCACC	GATACGACG	AGGGAAACGCC	GCCCGCCGGC	CGCTGCCCAT	TCGCGATCCA
	71521	GGACGGTCAC	CGCGCCATCC	TGGAGAGCGG	CACGGTGGGT	TCGTTCGACC	TGTTCGCGT
50	71581	CAAGCACTGG	CTGGTCGCCG	CCGCCGAGGA	CGTCAAGCTG	GTCACCAACG	ATCCGCGGTT
	71641	CAGCTCGGCC	GCGCCGTCCG	AGATGCTGCC	CGACCGCGG	CCCGGCTGGT	TCTCCGGGAT
	71701	GGACTCACCG	GAGCACAAAC	GCTACCGGCA	GAAGATCGCG	GGGGACTTCA	CACTGCGCGC
	71761	GGCGCGCAAG	CGGGAGGACT	TCGTCGCCGA	GGCCGCCGAC	GCCTGCCCTGG	ACGACATCGA
	71821	GGCCGCGGGGA	CCCGGCACCG	ACCTCATCCC	CGGGTACGCC	AAGCGGCTGC	CCTCCCTCGT
55	71881	CATCAACCGC	CTGTACGGGC	TCACCCCTGA	GGAGGGGCC	GTGCTGGAGG	CACGGATGCG
	71941	CGACATCACC	GGCTCGGCCG	ATCTGGACAG	CGTCAAGACG	CTGACCGACG	ACTTCTCGG
	72001	GCACGCGCTG	CGGCTGGTCC	GCGCGAAGCG	TGACGAGCGG	GGCGAGGAC	TGCTGCACCG
	72061	GCTGGCCTCG	GCCGACGACG	GCGAGATCTC	GCTCAGCGAC	GACGGAGCGA	CGGGCGTGT
	72121	CGCGACGCTG	CTGTTCGCCG	GCCACGACTC	GGTGCAGCGAG	ATGGTCGGCT	ACTGCCTCTA
60	72181	CGCACTGCTC	AGCCACCCCG	AGCAGCAGGC	GGCGCTGCGC	GCAGCGCCCGG	AGCTGGTCGA

72241 CAACCGGGTC GAGGGAGATGC TCCGTTTCCCT GCCCCGTCAAC CAGATGGGCG TACCGCGCGT
 72301 CTGTGTCGAG GACGTCGATG TGCGGGCGT GCGCATCCGT GCAGGGCGACA ACGTGATCCC
 72361 GCTCTACTCG ACGGCCAACCGA GCGACCCCGA GGTGTTCCCG CAGCCCGACA CCTTCGATGT
 72421 GACGCCCG CGTGGAGGGCA ACTTCGCGTT CGGCCACCGC ATTACAAGT GTCCCGGCCA
 5 72481 GCACATCGCC CGGGTGCTCA TCAAGGTGCG CTGCGCTGCGG TTGTTGAGC GTTCCCGGA
 72541 CGTCCGGCTG GCGGGCGACG TGCGGATGAA CGAGGGGCTC GGGCTGTTCA GCCCGGCCGA
 72601 GCTGCGGGTC ACCTGGGGGG CGGCATGAGT CACCCGGTGG AGACGTTGCG GTTGCCAAC
 72661 GGGACGACGG TCGCGCACAT CAACCGGGC GAGGCGCAGT TCCTCTACCG GGAGATCTC
 72721 ACCCAGCGCT GCTACCTGCG CCACCGTGTC GACCTGCGCC CGGGGGACGT GGTGTTGAC
 10 72781 GTCGGCGCGA ACATCGGCAT GTTCACGCTT TTGCGCGCATC TGGAGTGTCC TGGTGTGACC
 72841 GTGACGCCT TCGAGGCCG GCGCGTCCG TTGCGCGCGC TGCGGGCGAA CGTGACGCGG
 72901 CACGGCATCC CGGGCCAGGC GGACCGATGCG GCGGTCTCCG ACAGCTCCGG CACCCGGAAG
 72961 ATGACCTTCT ATCCCGACGC CACCGTGATG TCCGGTTTCC ACGCGGATGC CGCGGCCCGG
 15 73021 ACGGAGCTGT TGCGCACGCT CGGCCTCAAC GGCAGCTACA CGCCCGAGGA CGTCGACACC
 73081 ATGCTCGCGC AACTGCCCCA CGTCAGCGAG GAGATCGAAA CCCCTGTGGT CCGGCTCTCC
 73141 GACGTATCG CGGAGCGCGG TATCGAGGCC ATCGGCTCGC TGAAGGTCGA CGTGGAGAAC
 73201 AGCGAACGGC AGGTCTTCG CGGCCTCGAG GACACCGACT GGCCCCGTAT CCGCCAGGTC
 73261 GTCGCGGAGG TCCACGACAT CGACGGCGCG CTCGAGGAGG TCGTCACGCT GCTCCGCGC
 20 73321 CATGGCTTCA CCGTGGTCCG CGAGCAGGAA CGCCTGTTCG CCGGCACCGG CATCCACCAAG
 73381 GTCGCCGCGC GGCGGGTGGC CGGCTGAGCG CGCTCGGGC CGGGCCGCTC CGCACCGGGC
 73441 GCCGCGGTGC GGACGGCGGC TCAGCGGGCG TCGGACAGTT CCTGGGAGG TTGCTGACGG
 73501 CCCTCACCC CCAGCTTGC GAACACGTTG GTGAGGTGCT GTTCCACCGT GCTGGAGGTG
 73561 ACGAACAGCT GGCTGGCGAT CTCCCTGTTG GTGCGCCCGA CGCGGGCGTG CGACGCCACC
 73621 CGCCGCTCCG CCTCGGTCA CGATGTGATC CGCTCGCCG CGTCACGTC CTGGGTGCCG
 25 73681 TCCGCGTCCG AGGACTCCCC ACCGAGCCGC CGGAGGAGCG GCACGGCTCC GCACGGGTC
 73741 GCGAGGTGCC GTGCGCGCGC GAACAGTCCC CGCGCACCGC TGCGCGCCG GAGCATGCC
 73801 CACGCTCGC CCATGTCGGC GAGGACGCGG GCCAGCTCGT ACTGGTCGCG GCACATGATG
 73861 AGCAGATCGG CGGCCTCGTC GAGCAGTTG ATCCGCTTGG CGCGCGGACT GTAGGCCGCC
 73921 TGCACCCGCA GCGTCATCAC CGCGCCCGG GACCCCATCG CGCGGGACAG CTGCTCGGAG
 30 73981 ATGAGCCTCA GCCCCCTCGTC ACGGCCGCGG CGAGCAGCA GAAGCGCTTC GGCAGCGTC
 74041 ACCCGCCACA GGGCCAGGCC CGGCACGTCG ACGGACCAAG GTCGCATCCG CTCCCCCGAG
 74101 TCCCGGAACG CGTTGTACGC CGCCCGGTAC CGCCCGGGCG CGAGATGGTG TTGCCCACGG
 74161 GCCCAGACCA TGTGCACTCC GAAGAGGCTG TCGGAGGTCT CCTCCGGCAA CGGCTCGGCG
 74221 AGCCACCGCT CGGCCCGGT CAGGTCGCCC AGTCGGATCG CGCGGGCCAC GGTGCTGCTC
 35 74281 AGCGGCAATG CGGCGGCCAT CCCCCAGGAG GGCACGACCC GGGGGGGCGAG CGCGGCCCTCG
 74341 CCGCATTCGA CGGCGGGCGGT CAGGTCGCCC CGGCAGCGCG CGGCCTCGGC CGGGAACCCCC
 74401 GCGTGGACCG CCTCGTCGGC CGGGGTCCGC ATGTTGTCTG CACCCGGCAG CTTGTCGACC
 74461 CAGGACTTGA CGGCATCGGT GTCCTCGCG TAGAGCAGGG CCAGCAACGC CATCATGGTC
 74521 GTGGTCCGGT CCGTCGTGAC CGGGGAGTGC TGGAGCACGT ACTCGGCTTT GGCCTCGGCC
 40 74581 TGTTGCGACC AGCCGCGCAG CGCGTTGCTC AGGGCCTTGT CGCGACGGC GCGGTGCCGG
 74641 ACGGCTCCGG AAAACGAGGC GACCTCGTEC TCGGCCGGGG GATCGGCCGG ACGCGGCCGA
 74701 TCGGCCGCGC CGGGATAGAT CAGCGCGAGG GACAGGTCCG CGACGCGCAG GTGCGCCCGG
 74761 CCCTGCTCGC TCGGGGCGGC GGAGCGCTGG CGCGCCAGGA CCTCGGCCGG CTCGCCCCGGC
 74821 CGCCCGTCCA TCGCCAGCCA GCAGGGCAGC GACACGGCGT GCTCGCTGGA GAGGAGCCGT
 45 74881 TCCCGCGACG CGGTGAGCAG CTCGGCACA TGCCGGCCGG ATCTGGCGGG ATCGCAGAGC
 74941 CGCTCGATGG CGGCGGTGTC GACGCGCAGT CGGGCGTGA CGGGGGGTC GTCGGAGGCC
 75001 CGGTAGGCAG ACTCCAGGTA GGTGACGGCC TCGTCGAGCT CGCCGCGCAG GTGGTGTCTCG
 75061 CGCGCGCGT CGGTGAACAG CCCGGCGACC TCGGCCCGT GCACCCGGCC GGTACCCATC
 75121 TGGTGGCGGG CGAGCACCTT GCTGGCACG CGCGGGTCCC GCAGCAGTTC CAGCGCCAGC
 50 75181 TCGTGCAGGC CACGCCGCTC GGCGGGGAG AGGTGCTGA GTACGACGGA CGGGGCCCGC
 75241 GGGTGCAGGG ACCGCCCTTC CCGCAGCAGC CGCCCCCTGA CCAGCTGTT GCAGGGCTGC
 75301 TCGACCGCCT CGGTGTCGAG CGCGGTCTC CGCTGGACGA GGGTGAGTTC GACACTCTCG
 75361 CCGAGCACGG CGGAAGCTCG GGCGACGCTC AGCGCGGGCG GGCGCAACG ATAGAGCGAC
 75421 CGGAGGTAGG CGAGCCGGTA CGCCCGCCCC GCGACCACTT CGAGGCACCC TGAGGTCCGT
 55 75481 GTCCGTGCCT CCCGGATGTC GTCGATCAGG CGTGGCCGA GGAGCAGGTT CGCGCCGGTC
 75541 CGCCGGAACG CCTGGGCCAC CACGTCGTCG TGCGCGTCT CGCCGAGGTG CGCGCGCACG
 75601 AGTCGGTGG TCTGCGCTC GGTGAGCGGG CGCAGCGCGA TCTCTGTTA GTGGCGCAGA
 75661 CTCAGCAGTG CCGCCCGGAA TTGGGAGTGG CGGGGCGTCG GCCGGAGCAG CTCGGTCAGC
 75721 ACGATGGCGA CACGGGCCCG GCTGATGCGG CGCGCGAGGT GGAGCAGGCA CGCGAGCGAC
 60 75781 GGCGCGTCGG CGTGGTGCAC GTGCGTGCAT CGGATCAGTA CGGGCGCTC CGCGCGAGC

75841 GTCAGCACCG TGCGGGTGAG TTCGGTCCCC AGGCGGTTGT CGACGTCGGC CGGCAGGTT
 75901 TCGCACGATG CCGTCAGCCG GACCAGCTCC GGTGTCCGGG CGGCCAGCTC GGGCTGGTCG
 75961 AGGAGCTGGC CGAGCATGCC GTACGGCAGG GCCCCTCCT CCATGGAGCA CACCGCGCGA
 76021 AGGGTGACGA AGCCGGCTT GGCCGCGGCC GCGTCGAGGA GTTCGGTCTT GCGCAGGCG
 5 76081 ATCGGCCCCGG TGACGGCGGC GACGACGCC CGCCCGCCCC CCGCTCGGGT GAGCGCCCGG
 76141 TGGAGGGAAC CGAACTCGTC ATCGCGGGCG ATCAGGTCTG GGGGAGATAA GCGCCTATC
 76201 ACGAATGGAA CTACCTCGCG ACCGTCGTGG AAACCCATAG GCATCACATG GCTTGTGAT
 76261 CTGTACGGCT GTGATTCAAG CTGGCGGGAT GCTGTGCTAC AGATGGGAAG ATGTGATCTA
 76321 GGGCGTGCCT GTTCCCTCAG GAGCCGACCG CCCCCGGCGC CACCCGCCGT ACCCCCTGGG
 10 76381 CCACCAGCTC GGCGACCCGC TCCTGGTGGT CGACGAGGTA GAAGTGGCCG CGGGGAGAAGA
 76441 CCTCCACCGT GGTGGCGCG GTCGTGTGCC CGGCCAGGC GTGGGCCTGC TCCACCGTCG
 76501 TCTTCGGATC GTCGTCACCG ATGCACACCG TGATCGGCCT CGCCAGCGGC GGCAGGGCT
 76561 CCCACCGGTA CGTCTCCGCC GCGTAGTAGT CCGCCCGCAA CGGCGCCAGG ATCAGCGCG
 15 76621 GCATTCGTC GTCCGCCATC ACATCGCGC TCGTCCCGC GAGGCGCATG ACCGCCGCCA
 76681 GCAGCTCGTC GTCGGACCGC AGGTGGTCTT GGTCGGCGCG CGGCTGCGAC GGCAGGGCC
 76741 GGCCCGAGAC GATCAGGTGC GCCACCGGGA GCGCTGGGC CAGCTCGAAC GCGAGTGTG
 76801 CGCCCAGTGT GTGGCCGAAC AGCACCAAGCG GACGGTCCAG CCCCCGGCTTC AACGCTCTGG
 76861 CCACGAGGCC GGCGAGAACCA CGCAGGTGCGC GCACCGCCTC CTCGTCGCGG CGGTCTGGC
 20 76921 GGCCGGGGTA CTGCACGGCG TACACGTCCG CCACCGGGC GAGCGCACGG GCCAGGGAA
 76981 GGTAGAACGT CGCCGATCCG CGGGCGTGGG GCAGCAGCAC CACCCGTACCG GGGGCTCGG
 77041 GCGTGGGAA GAACTGCCG AGCCAGAGTT CCGAGCTCAC CGCACCCCCCT CGGCCGCGAC
 77101 CTGGGGAGCC CGGAACCGGG TGATCTGGC CAAGTGGCTTC TCCCGCATCT CGGGTGGT
 77161 CACGCCCAT CCCTCCTCCG GCGCCAGACA GAGGACGCCG ACTTTGCCGT TGTGACATT
 77221 GCGATGCACA TCGCGCACCG CCGACCCGAC GTCGTGAGC GGGTAGGTCA CCGACAGCGT
 25 77281 CGGGTGCACC ATCCCCTTGC AGATCAGGCG GTTCGCCTCC CACGCCCTCAC GATAAGTTCGC
 77341 GAAAGGGTA CCGATGATCC GCTTCACCGA CATCCACAGG TACCGATTGT CAAAGGCGTG
 77401 CTCGTATCCC GAGGTTGACG CGCAGGTGAC GATCGTGCCA CCCCCGACGTG TCACGTAGAC
 77461 ACTCGCGCCG AACGTGCGC GCCCCGGGTG CTCGAACACG ATGTCGGGAT CGTCACCGCC
 77521 GGTCAGCTCC CGGATC

30

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the FK-520 PKS of *Streptomyces hygroscopicus* is shown herein merely to illustrate a preferred embodiment of the invention, and the present invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences shown merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following general description of the FK-520 PKS genes and modules of the PKS proteins encoded thereby is provided. This general description is followed by a more detailed description of the various domains and modules of the FK-520

PKS contained in and encoded by the compounds of the invention. In this description, reference to a heterologous PKS refers to any PKS other than the FK-520 PKS. Unless otherwise indicated, reference to a PKS includes reference to a portion of a PKS. Moreover, reference to a domain, module, or PKS includes reference to the nucleic acids encoding the 5 same and vice-versa, because the methods and reagents of the invention provide or enable one to prepare proteins and the nucleic acids that encode them.

- The FK-520 PKS is composed of three proteins encoded by three genes designated *fkbA*, *fkbB*, and *fkbC*. The *fkbA* ORF encodes extender modules 7 - 10 of the PKS. The *fkbB* ORF encodes the loading module (the CoA ligase) and extender modules 1 - 4 of the PKS. 10 The *fkbC* ORF encodes extender modules 5 - 6 of the PKS. The *fkbP* ORF encodes the NRPS that attaches the pipecolic acid and cyclizes the FK-520 polyketide.

The loading module of the FK-520 PKS includes a CoA ligase, an ER domain, and an ACP domain. The starter building block or unit for FK-520 is believed to be a dihydroxycyclohexene carboxylic acid, which is derived from shikimate. The recombinant 15 DNA compounds of the invention that encode the loading module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of methods and in a variety of compounds. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding 20 sequence for the loading module of the heterologous PKS is replaced by the coding sequence for the FK-520 loading module, provides a novel PKS coding sequence. Examples of heterologous PKS coding sequences include the rapamycin, FK-506, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that 25 comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, either replacing the CoA ligase with a different CoA ligase, deleting 30 the ER, or replacing the ER with a different ER. In addition, or alternatively, the ACP can be replaced by another ACP. In similar fashion, the corresponding domains in another loading or extender module can be replaced by one or more domains of the FK-520 PKS. The resulting heterologous loading module coding sequence can be utilized in conjunction

with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide.

The first extender module of the FK-520 PKS includes a KS domain, an AT domain specific for methylmalonyl CoA, a DH domain, a KR domain, and an ACP domain. The 5 recombinant DNA compounds of the invention that encode the first extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for 10 a module of the heterologous PKS is either replaced by that for the first extender module of the FK-520 PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence 15 for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or only a portion of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2- 20 hydroxymalonyl CoA specific AT; deleting either the DH or KR or both; replacing the DH or KR or both with another DH or KR; and/or inserting an ER. In replacing or inserting KR, DH, and ER domains, it is often beneficial to replace the existing KR, DH, and ER domains with the complete set of domains desired from another module. Thus, if one desires to insert an ER domain, one may simply replace the existing KR and DH domains with a KR, DH, 25 and ER set of domains from a module containing such domains. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a gene for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting 30 heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous

PKS can be replaced by one or more domains of the first extender module of the FK-520 PKS.

In an illustrative embodiment of this aspect of the invention, the invention provides recombinant PKSs and recombinant DNA compounds and vectors that encode such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and domains of an FK-520 or FK-520 derivative PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS encoded thereby can be fed or supplied with N-acylcysteamine thioesters of novel precursor molecules to prepare FK-520 derivatives. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication Nos. US97/02358 and US99/03986, each of which is incorporated herein by reference.

The second extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the second extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 second extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the second extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these

replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous second extender module coding sequence 5 can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the second extender module of the FK-520 PKS.

The third extender module of the FK-520 PKS includes a KS, an AT specific for 10 malonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the third extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. 15 The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the FK-520 PKS is inserted into a DNA 20 compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 25 malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding 30 sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-520, an

FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the third extender module of the FK-520 PKS.

- The fourth extender module of the FK-520 PKS includes a KS, an AT that binds ethylmalonyl CoA, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the fourth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS.
- 10 The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the FK-520 PKS is inserted into a
- 15 DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the ethylmalonyl CoA specific AT with a malonyl CoA, methylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; and/or deleting the inactive DH, inserting a KR, a KR and an active DH, or a KR, an active DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, a PKS for a polyketide other than FK-520, or from chemical synthesis.

25 The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fourth extender module of the FK-520 PKS.

As illustrative examples, the present invention provides recombinant genes, vectors, and host cells that result from the conversion of the FK-506 PKS to an FK-520 PKS and vice-versa. In one embodiment, the invention provides a recombinant set of FK-506 PKS

genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of the fourth extender module of the FK-520 PKS. This recombinant PKS can be used to produce FK-520 in recombinant host cells. In another embodiment, the invention 5 provides a recombinant set of FK-520 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of the fourth extender module of the FK-506 PKS. This recombinant PKS can be used to produce FK-506 in recombinant host cells.

Other examples of hybrid PKS enzymes of the invention include those in which the 10 AT domain of module 4 has been replaced with a malonyl specific AT domain to provide a PKS that produces 21-desethyl-FK520 or with a methylmalonyl specific AT domain to provide a PKS that produces 21-desethyl-21-methyl-FK520. Another hybrid PKS of the invention is prepared by replacing the AT and inactive KR domain of FK-520 extender module 4 with a methylmalonyl specific AT and an active KR domain, such as, for 15 example, from module 2 of the DEBS or oleandolide PKS enzymes, to produce 21-desethyl-21-methyl-22-desoxo-22-hydroxy-FK520. The compounds produced by these hybrid PKS enzymes are neurotrophins.

The fifth extender module of the FK-520 PKS includes a KS, an AT that binds methylmalonyl CoA, a DH, a KR, and an ACP. The recombinant DNA compounds of the 20 invention that encode the fifth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS 25 is either replaced by that for the fifth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that 30 produces an FK-520 derivative.

In another embodiment, a portion of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA